

QUANTITATIVE ASSESSMENT OF
MMWAVE-BASED POINT CLOUD
GENERATION PIPELINES FOR TARGET
DETECTION

QUANTITATIVE ASSESSMENT OF MMWAVE POINT CLOUD
FOR TARGET DETECTION

By BOYU JIANG, B.Eng

A Thesis Submitted to the School of Graduate Studies in Partial
Fulfillment of the Requirements for
the Degree Master of Applied Science

McMaster University © Copyright by Boyu Jiang, December 2024

McMaster University

MASTER OF APPLIED SCIENCE (2024)

Hamilton, Ontario, Canada (Computing and Software)

TITLE: Quantitative Assessment of mmWave Point Cloud for
Target Detection

AUTHOR: Boyu Jiang
B.Eng (Software Engineering),
McMaster University, Hamilton, Canada

SUPERVISOR: Dr. Rong Zheng

NUMBER OF PAGES: xv, 89

Lay Abstract

This thesis explores ways to measure the accuracy and reliability of a 3D representation called a point cloud created from radar data. Radar, like the kind used in car driving assist features, can detect objects and people by transmitting high-frequency radio waves. This thesis focuses on developing evaluation methods to assess how closely point clouds match objects or people in the real world.

Experiments were conducted in a laboratory environment to validate these evaluation methods. Static corner reflectors, known for their strong radar signal reflection, were used to assess the system's performance in capturing exact positions. Additionally, the study examined the quality of point clouds in representing moving human subjects.

The objective is to develop a set of metrics that quantify the performance of various point cloud generation methods. These quantitative measures can improve the accuracy and reliability of the point cloud generation pipeline, thereby improving downstream applications such as indoor localization and monitoring.

Abstract

This thesis tackles the challenge of employing quantifiable metrics to assess the quality of point clouds generated by various distinct pipelines using TI IWR6843AOP mmWave FMCW radar. This study focuses on developing quantifiable metrics to evaluate point cloud quality for both static targets, such as the corner reflectors used in this research, and human targets.

For static point targets, this study introduces metrics that combine Euclidean distance errors with range, azimuth, and elevation angle errors, providing a more comprehensive assessment compared to using Euclidean distance errors alone.

For human targets, this thesis introduces metrics from two perspectives. The first focuses on coverage, employing the Euclidean distance to the human mesh surface to quantify errors between the ground truth human mesh and the point cloud. Additionally, It calculates the Euclidean distance between each point and all joints, selecting the minimum distance to determine the closest joint and evaluating the percentage of points reflected from each body segment. The second focus is on consistency. Point cloud consistency across consecutive frames is assessed by analyzing the mean and maximum intensity values and calculating Hausdorff distances to evaluate the stability of the point cloud distribution.

To my family and friends

Acknowledgements

There are many people I would like to thank for their support.

I would like to begin by expressing my gratitude to my supervisor, Dr. Rong Zheng, for her exceptional guidance and unwavering patience, both of which were crucial to the successful completion of this thesis.

I would also like to my heartfelt thanks to my partner Yujiao for her unwavering support during challenging times and her encouragement in the face of obstacles.

Additionally, I would like to express my gratitude to Zimeng Zhou and Yunkai Yu for their invaluable assistance in acquiring the ground truth data essential for this work. Their generosity with their time has been immensely appreciated.

Finally, I would like to extend my deepest gratitude to my family for their unwavering support, not only during the challenging times of completing this work but also throughout my years of pursuing higher education in Canada. They have been the cornerstone of my long journey, always just a phone call away.

Table of Contents

Lay Abstract	iii
Abstract	iv
Acknowledgements	vi
Notation, Definitions, and Abbreviations	xiii
Declaration of Academic Achievement	xvi
1 Introduction	1
1.1 Motivation	1
1.2 Organization of the thesis	6
2 Background	7
2.1 FMCW Radar Fundamental Principle	7
2.2 Texas Instrument Radar Characteristics	9
3 Point Cloud Generation	12
3.1 Clutter Removal	14
3.2 Spectrum Estimation	16

3.3	Target Detection	21
3.4	Spherical coordinates to Cartesian coordinates	30
4	Evaluation Metrics	32
4.1	Static Point Target Metric	32
4.2	Human Target Metric	35
5	Performance Evaluation	38
5.1	Testbed Setup	38
5.2	Ground Truth Generation	42
5.3	Static Point Target Test	45
5.4	Human Subject Test	63
6	Conclusion	81
6.1	Limitations and Future Work	82

List of Figures

1.1	Examples of different representation of radar data	3
2.1	Frequency of chirp signal as a function of time	8
2.2	Basis of angle of arrival estimation	10
2.3	Antenna array geometry of TI IWR6843AOP.	11
3.1	Representation of Radar Data Cube	13
3.2	Overview of pipeline	13
3.3	Flow chart of first-order MTI filter transfer function	15
3.4	Frequency Response of First-Order MTI Filter	16
3.5	Spatial Smoothing sub-arrays	22
3.6	CUT, Guard Cells and Training Cells.	24
5.1	Corner reflector marker placement	40
5.2	Lab space setup for single corner reflector	40
5.3	Three corner reflectors scenario setup	41
5.4	Human subject scenario setup	42
5.5	Radar and marker placement	43
5.6	Corner reflector marker placement	44
5.7	Motion capture suit and marker placement	44
5.8	An Example of SMPL Human Mesh	45

5.9	Bar Graph of Average Error across All Trials	50
5.10	Euclidean Distance Error of Single-target Test.	51
5.11	Range Error of Single-target Test.	51
5.12	Azimuth Angle Error of Single-target Test.	52
5.13	Elevation Angle Error of Single-target Test.	52
5.14	Bar Plot of Mean Errors Across Corner Reflectors and All Methods in the Three-Target Test Scenario	62
5.15	Histogram of Euclidean distances between points and the ground-truth human mesh in the Standing scenario	71
5.16	Histogram of Euclidean distances between points and the ground-truth human mesh for the Walking in place scenario	72
5.17	Histogram of Euclidean distances between points and the ground-truth human mesh in the Walking scenario	73
5.18	Histogram of Euclidean distances between points and the ground-truth human mesh in the Dancing scenario	74
5.19	Histogram of Euclidean distances between points and the ground-truth human mesh in the Sitting scenario	75

List of Tables

2.1	Specifications and setting of IWR6843AOP	11
5.1	Pipelines for Static Target test	46
5.2	CFAR settings for Static Target Test	46
5.3	Ground Truth of Trials for Single-target Test	47
5.4	Mean and Standard Deviation of Trial 1 for all methods in single-target static test	53
5.5	Mean and Standard Deviation of Trial 2 for all methods in single-target static test	54
5.6	Mean and Standard Deviation of Trial 3 for all methods in single-target static test	55
5.7	Mean and Standard Deviation of Trial 4 for all methods in single-target static test	56
5.8	Mean and Standard Deviation of Trial 5 for all methods in single-target static test	56
5.9	Mean and Standard Deviation of Trial 6 for all methods in single-target static test	57
5.10	Mean and Standard Deviation of Trial 7 for all methods in single-target static test	58

5.11 Mean and Standard Deviation of Trial 8 for all methods in single-target static test	59
5.12 Mean and Standard Deviation of Trial 9 for all methods in single-target static test	60
5.13 Mean and Standard Deviation of Trial 10 for all methods in single- target static test	60
5.14 Ground Truth of Trials for Three-target Test	61
5.15 Pipelines for Human Subject test	64
5.16 Different CFAR parameter settings	65
5.17 the proportion of points identified as reflections from human body for different parameter settings	65
5.18 Hausdorff of points identified as reflections from human body for dif- ferent parameter settings	66
5.19 proportion of points identified as reflections from human body	67
5.20 Average number of points in point clouds across all frames	67
5.21 Coverage for body segments	69
5.22 Hausdorff Distance of points classified as reflection from human body	76
5.23 Hausdorff Distance of all points in point clouds	76
5.24 Standard Deviation of Mean Intensities across all frames	77
5.25 Standard Deviation of Maximum Intensities across all frames	78
5.26 Statistics for Maximum intensities of standing scenario	79
5.27 Statistics for Mean intensities of standing scenario	79
5.28 Statistics for Maximum intensities of sitting scenario	80
5.29 Statistics for Mean intensities of sitting scenario	80

Notation, Definitions, and Abbreviations

Notation

$\eta(\mathbf{p})$	Intensity of point \mathbf{p}
$\text{Med}(\mathbf{v})$	Median value of vector \mathbf{v}
$\arg \min \mathbf{v}$	Index of minimum value of vector \mathbf{v}
$d(a, b)$	Shortest Euclidean distance between a and b
$d_H(A, B)$	Hausdorff distance between point cloud A and B

Definitions

Target detection

The process of locating the object of interest within a scene. Target detection algorithms analyze the received data to detect the target object.

Point cloud A point cloud is a representation of radar data, consisting of selected points with their corresponding three-dimensional Cartesian coordinates, and potentially include other attributes, such as intensity of points.

Pipeline A sequential data processing pipeline designed to transform raw radar signals into high-level outputs, such as point clouds. Each stage in the pipeline involves modular components with a range of algorithmic options, enabling processing and extraction of information from the input data.

Evaluation metrics

A set of quantitative metrics designed to evaluate the performance of point cloud generation pipelines. These metrics provide detailed assessments of how effectively the generated point cloud captures the geometric and structural properties of the scene, as well as the reliability of target detection algorithms in identifying the target object.

Abbreviations

FMCW Frequency-Modulated Continuous-Wave

MIMO Multiple-input multiple-output

mmWave Millimeter wave

ADC Analog-to-digital converter

FOV	Field of View
AOA	Angle of arrival
FFT	Fast Fourier transform
FTI	Finite impulse response
MTI	Moving target indication
MVDR	Minimum Variance Distortion-less Response
MUSIC	Multiple signal classification
CFAR	Constant false alarm rate
CDF	Cumulative distribution function

Declaration of Academic Achievement

I, Boyu Jiang, hereby declare that this thesis is my original work and that I am its sole author. No portion of this thesis has been submitted to fulfill the requirements for a degree at any other institution.

To the best of my knowledge, the content of this thesis does not violate the copyright of any third party.

My supervisor, Dr. Rong Zheng, has provided invaluable support throughout every stage of this work. Additionally, my colleagues, Zimeng Zhou and Yunkai Yu, assisted in the acquisition of ground truth data. All research activities were conducted independently by me.

Chapter 1

Introduction

1.1 Motivation

Remote sensing using mmWave radar has become increasingly popular in many fields, such as smart homes and in-home health monitoring. These fields encompass a range of tasks but are not limited to vital sign monitoring [30], people counting [10, 19], posture classification [22], fall detection [32], as well as human target tracking and classification [28] and construction [29, 31]. Various forms of radar data representations are explored for different kinds of tasks, such as range maps, range angle maps, 2D heat maps in Cartesian coordinates and point clouds.

The choice of representation depends on the task. Range map as shown in Figure 1.1a has combined information on every range bin, it does not contain any angular information. For tasks like single-person vital sign monitoring, only the range map is used to identify the range bin containing the person because there is no need to detect a second person or object, tracking the range bin for each frame is sufficient for further vital sign extraction [30].

The range-angle map is a data representation where one axis corresponds to range, and the other axis represents either azimuth or elevation angles within a single frame as demonstrated in Figure 1.1b. For object detection and classification, where three-dimensional positional information is unnecessary but having information on reflection intensities and object dimensions is essential, the range-angle map provides a suitable data representation. For instance, a car and a human vary in the number range bin and angle cells each occupies [8].

Human representation can use a 2D heat map in Cartesian coordinates [33]. A 2D heat map is a projection derived from a three-dimensional Cartesian heat map with x, y, and z axes. The 2D heat map is generated by reducing one axis, typically using either the average or maximum value along that axis, leaving only two axes for visualization as illustrated in Figure 1.1c. However, for tasks requiring three-dimensional positional data, such as localization and fall detection, a more complex form of data representation is needed such as point clouds [32, 14].

A point cloud as shown in Figure 1.1d is a representation of radar data, consisting of selected points with their corresponding three-dimensional Cartesian coordinates, intensity values, and optionally velocity information. The Cartesian coordinates are derived from the azimuth and elevation angles obtained from the radar antenna array. Point clouds provide more comprehensive information compared to other forms of data representations. Compared to range-angle maps and 2D Cartesian heat maps, three-dimensional point clouds offer a more comprehensive representation of spatial information. Range-angle maps and 2D heatmaps can provide only partial insights into the environment as they are inherently limited. Additionally, these representations are more sensitive to the placement and orientation of the radar system induced

by their two-dimensional nature. In contrast, three-dimensional point clouds capture the full spatial geometry of the scene, enabling coordinate transformations such as coordinate rotations. This makes them particularly suitable for applications like 3D indoor human tracking, where precise position and height of the person relative to the room, rather than the device placement, are essential for accurate analysis.

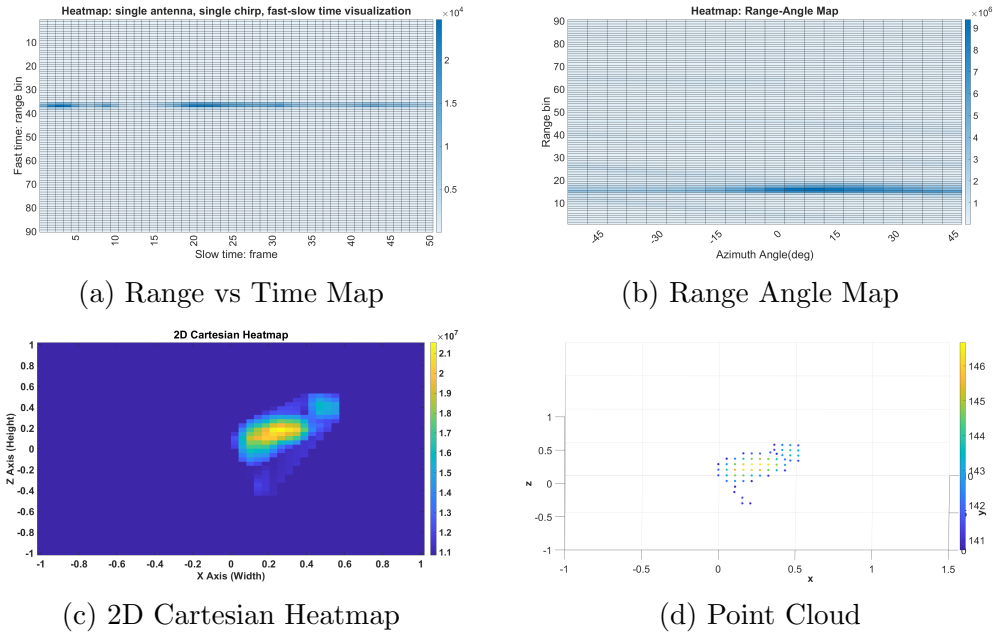


Figure 1.1: Examples of different representation of radar data

This research addresses two research questions.

- What are the objective measures to quantify the quality of point clouds for target detection.
- How various pre-processing pipelines and different parameter configurations for point cloud generation perform when assessed using these metrics.

For tasks that utilize point clouds as inputs, the quality of the point cloud will impact the performance of downstream processes, regardless of whether conventional point cloud processing or neural networks are applied. Therefore, **quantitative metrics** are essential for evaluating point clouds, enabling a thorough understanding of their quality and identifying potential errors. The conventional method that utilizes Euclidean distance between the ground truth point cloud and captured or simulated point cloud could provide information on how closely the point cloud generated resembles the ground truth. In [5], the mean Euclidean distance is calculated between the reconstructed point cloud and key ground-truth landmarks, similarly in [18], it is computed between the simulated and ground-truth point clouds. Euclidean distance serves as an effective metric for quantifying localization accuracy for discrete point targets. However, mean Euclidean distance alone does not account for the spatial coverage of the point cloud relative to the full extent of the human target, limiting its effectiveness in assessing comprehensive representation. In addition, Euclidean distance lacks the capacity to provide insight into the underlying factors contributing to the observed error.

Another metric commonly employed is Intersection over Union (IoU), it provides information on how much ground truth and point cloud overlap. In [6], IoU metrics are utilized to compare the bounding box of the ground truth with the predicted bounding box derived from point clouds. In [20], IOU metrics are calculated based on the predicted point cloud and the true point cloud derived from the groundtruth. IoU metric quantifies how much the generated point cloud overlaps with the groundtruth point cloud. However, this method does not account for the coverage of the point cloud, which is influenced by factors such as the observed object’s geometry, the

radar’s placement, and its Field of View (FOV). The IoU metric based on bounding boxes is primarily suitable for objects with rigid shapes. For human subjects with complex and articulated body structures, bounding box based IoU can overestimate the occupied space by including background regions, leading to inaccurate localization and assessment of detection accuracy. As demonstrated in [24], the distribution of reflection points from targets such as the human body is inherently non-uniform and varies depending on the subject’s posture and the radar’s positioning. However, the ground-truth-derived point cloud uniformly represents the entire human body, which can lead to discrepancies in the IoU metric assessment due to mismatched spatial coverage.

Neither Euclidean distance nor IoU provides insight into the underlying sources of error. Both metrics are limited in quantifying point cloud coverage of the human body, as they primarily focus on spatial distribution while neglecting temporal consistency. To overcome these limitations, this study introduces refined quantitative metrics and evaluates the performance of different point cloud generation pipelines for both human and static targets.

- For static targets, this work proposes a metric composed of Euclidean distance error, range error, and azimuth and elevation angle errors, providing a more comprehensive assessment of point cloud quality compared to relying solely on Euclidean distance. Five distinct pipelines are evaluated, with the pipeline incorporating MVDR for angle estimation and the variance-based target detection algorithm demonstrating the highest overall performance.
- For human targets, this work introduces a combination of coverage evaluation and consistency evaluation. Coverage assessment measures the proportion of

points originating from human reflections and evaluates the extent of coverage across different body parts under various scenarios. Consistency evaluation involves analyzing the intensity of reflections across frames and computing the Hausdorff distance between frames, assuming the human target remains stationary with no changes in its 3D Cartesian position relative to the radar. The pipeline involving MVDR as spectrum estimation, Mean removal as clutter removal, and CFAR as target detection has the best overall performance. An alternative pipeline employing an MTI filter for clutter removal and the variance-based method for target detection exhibited comparable performance in coverage assessment across most scenarios.

1.2 Organization of the thesis

Chapter 2 introduces the fundamental principles of FMCW radar and highlights key characteristics of the radar utilized in this study. Chapter 3 will cover the pipeline of radar point cloud generations and various algorithms employed. Chapter 4 will discuss the quantitative evaluation metrics. Chapter 5 outlines the experimental setup and ground truth generation process, followed by a presentation of the evaluation results.

Chapter 2

Background

2.1 FMCW Radar Fundamental Principle

A Frequency-Modulated Continuous-Wave (FMCW) radar transmits signals at regular intervals, a chirp $s(t)$ transmitted at time t is defined as

$$s(t) = A_t e^{j(2\pi f_0 t + \pi S t^2)}, \quad 0 \leq t \leq T_s. \quad (2.1.1)$$

The chirp signal $s(t)$ has its frequency linearly increasing over the period T_s from the starting frequency f_0 . A_t is the complex amplitude of the signal. The slope S of the linear frequency modulation is defined as $\frac{(f_{T_s} - f_0)}{T_s}$, where f_{T_s} represents the maximum frequency attained at the end of the period, and bandwidth $B = f_{T_s} - f_0$.

For signal reflected from a single point target, the received signal $r(t)$ has a time delay $t_d(t)$ to the transmitted signal $s(t)$.

$$r(t) = A_t A_r e^{j(2\pi f_0(t-t_d(t)) + \pi S(t-t_d(t))^2)}, \quad 0 \leq t \leq T_s. \quad (2.1.2)$$

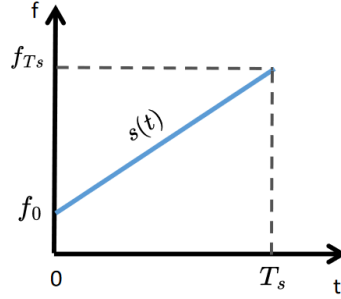


Figure 2.1: Frequency of chirp signal as a function of time

The time delay $t_d(t)$ is given by $\frac{2R(t)}{c}$ [2], where $R(t)$ represents the range between the radar and the target at time t . c denotes the constant for the speed of light. $r(t)$ and $s(t)$ are combined via a mixer to produce an intermediate frequency (IF) signal, $y(t)$ is

$$y(t) = A_t A_r e^{j2\pi f_0 t - 2\pi f_0 (t - t_d(t)) + \pi S t^2 - \pi S (t - t_d(t))^2} \quad 0 \leq t \leq T_s, \quad (2.1.3)$$

The equation above can be reformulated as follows:

$$y(t) = A_t A_r e^{j(4\pi \frac{r(t)}{\lambda_{max}} + 4\pi S \frac{r(t)}{c} t - 4\pi S \frac{r^2(t)}{c^2})}, \quad 0 \leq t \leq T_s, \quad (2.1.4)$$

where $\lambda_{max} = \frac{c}{f_0}$. In the case of a stationary target, $y(t)$ is, in fact, a pure tone signal, for two stationary objects with a range difference of Δr , the difference in their frequencies is given by $\Delta f = \frac{2S\Delta r}{c}$. Since the frequency resolution is given by $\Delta f > \frac{1}{T_s}$, we have $\frac{2S\Delta r}{c} > \frac{1}{T_s}$, or equivalently $\Delta r = \frac{c}{2ST_s} = \frac{c}{2B}$ [16]. It demonstrates that range resolution Δr only depends on the bandwidth B of the chirp transmitted.

FMCW radar can also transmit multiple chirps continuously within one frame, (2.1.4)

can be discretized as

$$y(i, m) = A_t A_r(t_m) e^{j(4\pi \frac{r(m)}{\lambda_{max}} + 4\pi S \frac{r(m)}{c} i)}, \quad 1 \leq i \leq n_k, \quad 1 \leq m \leq n_c. \quad (2.1.5)$$

The third term in the exponent in (2.1.4) is omitted due to its negligible magnitude in indoor sensing, when the range $r(t)$ is small. n_k represents the number of fast-time samples within a chirp signal, corresponding to the number of range bins. The second term in (2.1.5) $e^{j(4\pi S \frac{r(m)}{c} t_i)}$ can be translated to range information. The first term represents the phase for each range bin. n_C denotes the number of slow-time samples, or equivalently, the number of transmitted chirps. Multiple chirps at the same range bin can be used to extract Doppler information. To avoid ambiguity, the phase shift between two consecutive chirps must remain below π , requiring that $\frac{4\pi v T_s}{\lambda_{max}} < \pi$. Hence, the maximum relative speed v_{max} can be measured by two consecutive chirps is [16]

$$v_{max} = \frac{\lambda_{max}}{4T_s}. \quad (2.1.6)$$

2.2 Texas Instrument Radar Characteristics

The radar used for this work is Texas Instrument IWR6843AOP. TI IWR6843AOP is an FMCW multiple-input, multiple-output (MIMO) radar. It has 4 receiving and 3 transmitting antennas. The multiple antenna configuration enables angle estimation. Angle estimation is induced by the physical separations between antennas, as illustrated in the figure below.

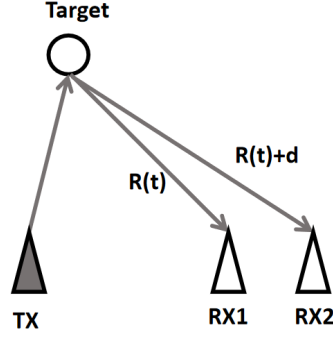


Figure 2.2: Basis of angle of arrival estimation

Figure 2.2 shows that for the same chirp and the same range bin, the target travels an additional distance d to RX2 than RX1 due to the geometry of the antenna array. In practice d is usually small, introducing an extra phase to the first term in (2.1.5), expressed as $e^{j(4\pi \frac{r(t_m)+d}{\lambda_{max}})}$. This additional phase term can be further exploited to estimate the angle of arrival (AoA) of targets. The TI IWR6843AOP FMCW MIMO radar, equipped with four receive antennas and three transmit antennas, forms a virtual array of 12 elements. This configuration enables the radar to achieve a 120-degree azimuth field of view (FOV) and a 120-degree elevation FOV [13]. The antenna geometry is illustrated in figure 2.3.

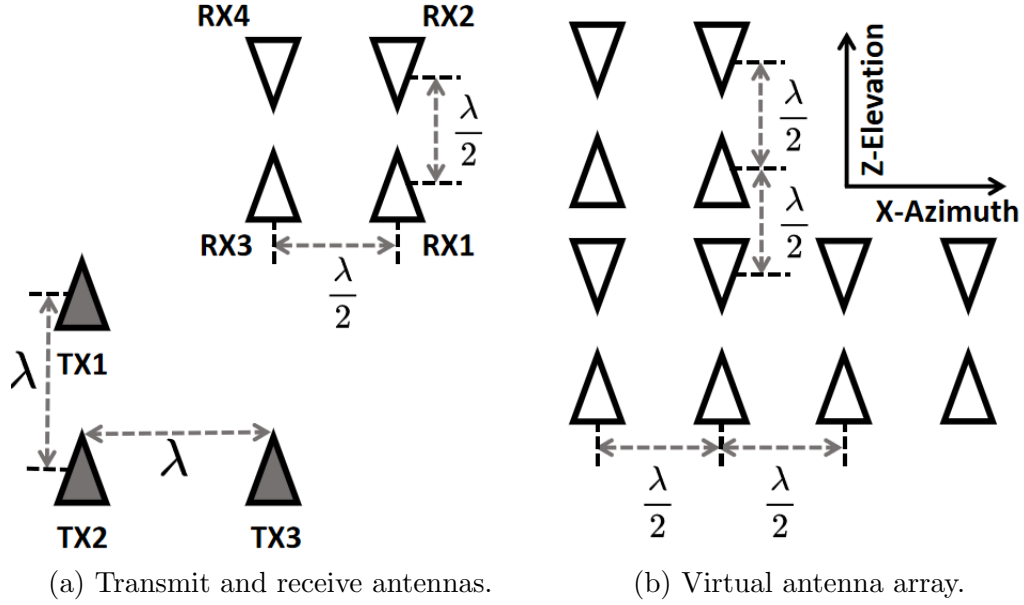


Figure 2.3: Antenna array geometry of TI IWR6843AOP.

As shown in Figure 2.3, spacing between virtual antennas is $\frac{\lambda}{2}$, where $\lambda \approx 5\text{mm}$. TI IWR6843AOP radar is capable of transmitting chirps at frequencies between 60GHz to 64GHz [13]. The specifications and important setting for this work is shown in Table 2.1.

Chirp Duration	ADC Sample Rate	Number of ADC Samples
50 μs	4500 kps	225
Range Resolution	Elevation Angular Resolution	Azimuth Angular Resolution
4.41 cm	29 degree	29 degree
Elevation FOV	Azimuth FOV	Frequency Start
+/- 60 degree	+/- 60 degree	60GHz
Frequency End	Frequency Slope	
63.4GHz	68MHz/ μs	

Table 2.1: Specifications and setting of IWR6843AOP

Chapter 3

Point Cloud Generation

The raw **ADC** (analog-to-digital converter) data is initially partitioned based on the number of frames n_f . For each frame, f , the raw data is organized into a three-dimensional complex-valued data cube with dimensions (v, n_k, n_c) as demonstrated in Figure 3.1, where v is the number of virtual antenna channels. n_k represents the number of fast-time sample points, corresponding to the number of range bins after range estimation. n_c denotes the number of chirps or the slow time samples transmitted per frame. Consistently, all frames within a single trial have an identical number of transmitted chirps and sampled points. As shown in (2.1.4), differences in range result in frequency variations within the raw data. To extract range information from the raw ADC data of a single frame, a discrete Fourier transform must be applied along the fast-time sample axis.

$$Y[k] = \sum_{t_i=1}^{n_k} y[t_i] e^{j \frac{(-2\pi)(t_i-1)(k-1)}{n_k}}, 1 \leq k \leq n_k. \quad (3.0.1)$$

This Fourier Transformed applied is the range estimation of FMCW radar. The

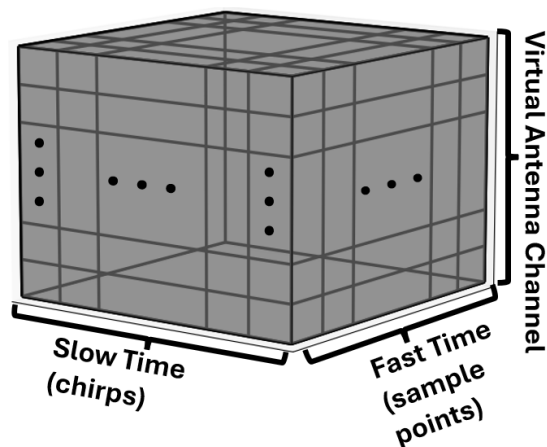


Figure 3.1: Representation of Radar Data Cube

new array Y is referred to as the range map. Range estimation is the first step of FMCW radar data processing. Various pipelines employing different methods have been proposed in numerous other studies. In this work, a generic pipeline comprising range estimation, clutter removal, spectrum estimation, target detection, and point generation is proposed to facilitate understanding.

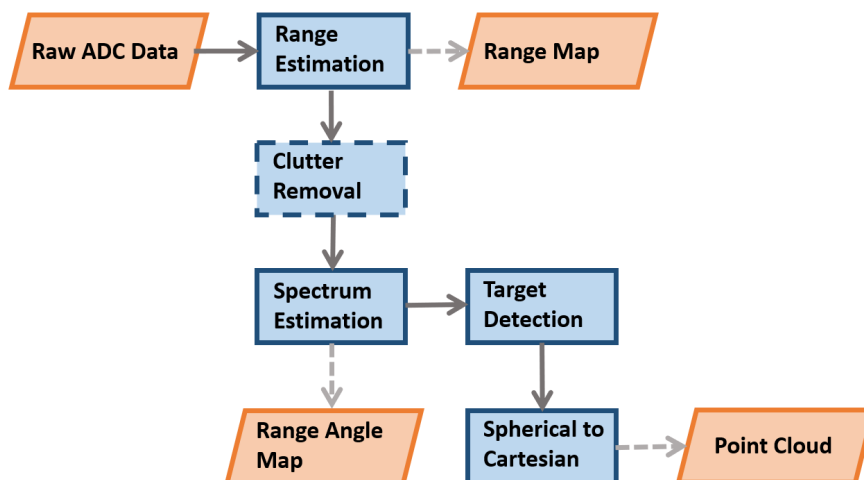


Figure 3.2: Overview of pipeline

As illustrated in the figure above, the clutter removal, spectrum estimation, and

target detection modules follow the range estimation module. Clutter removal is designed to remove static clutter from the signal, though it is typically not applied when the target of interest is stationary. Spectrum estimation, as the name suggests, determines the **angle of arrival (AOA)** of the target. Target detection employs methods to select points of interest that will be included in the point cloud. The point cloud is generated by converting the spherical coordinates of the points of interest from the radar into Cartesian coordinates. In the Clutter Removal, Angle Estimation, and Target Detection modules, each consists of multiple algorithmic implementations tailored to their respective functions. This chapter will examine some commonly employed methods for each module introduced above.

3.1 Clutter Removal

3.1.1 Mean Removal

One of the most popular techniques for clutter removal is mean removal. As demonstrated in (2.1.5), each frame consists of n_c number of chirps. To eliminate clutter, the average signal for each range bin is initially computed across the slow-time dimension.

$$Y_{clutter}[i] = \frac{1}{n_c} \sum_{k=1}^{n_c} Y[i, k], 1 \leq k \leq n_c, \quad (3.1.1)$$

The computed clutter signal is subtracted from all chirps in the same range bin, resulting in a clutter-free range map.

$$\hat{Y}[i, k] = Y[i, k] - Y_{clutter}[i], 1 \leq i \leq n_k. \quad (3.1.2)$$

Two variables that influence the performance of this approach are the number of chirps n_c , and the time interval between consecutive chirp signals. However, in this study, both variables remain constant across all trials.

3.1.2 FIR Filter for MTI

Another approach to static clutter removal is Moving Target Indication (MTI) filtering. The first-order MTI filter operates by subtracting the signal response of one chirp from the subsequent chirp [1].

$$\hat{Y}[i, k - 1] = Y[i, k] - Y[i, k - 1], 1 \leq i \leq n_k, 2 \leq k \leq n_c, \quad (3.1.3)$$

where n_k and n_c are the number of range bins and the number of chirps. (3.1.3) is also called two-pulse MTI cancellers. The transfer function of this filter is simply $H(z) = 1 - z^{-1}$, where $z = e^{j2\pi f_d T_s}$. f_d is the analog frequency. $\pi f_d T$ denotes in normalized radian frequency $\frac{\hat{\omega}}{2}$. The transfer function can be rewritten as

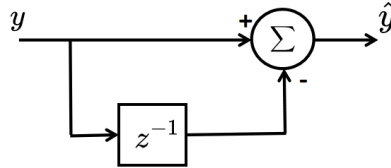


Figure 3.3: Flow chart of first-order MTI filter transfer function

$$\begin{aligned}
 H(\hat{\omega}) &= 1 - e^{-j\hat{\omega}} \\
 &= e^{-j\frac{\hat{\omega}}{2}} (e^{j\frac{\hat{\omega}}{2}} - e^{-j\frac{\hat{\omega}}{2}}) \\
 &= 2je^{-j\frac{\hat{\omega}}{2}} \sin\left(\frac{\hat{\omega}}{2}\right).
 \end{aligned} \quad (3.1.4)$$

The transfer function response of the first-order MTI filter provides significant attenuation for the zero-frequency components, as illustrated in Figure 3.4.

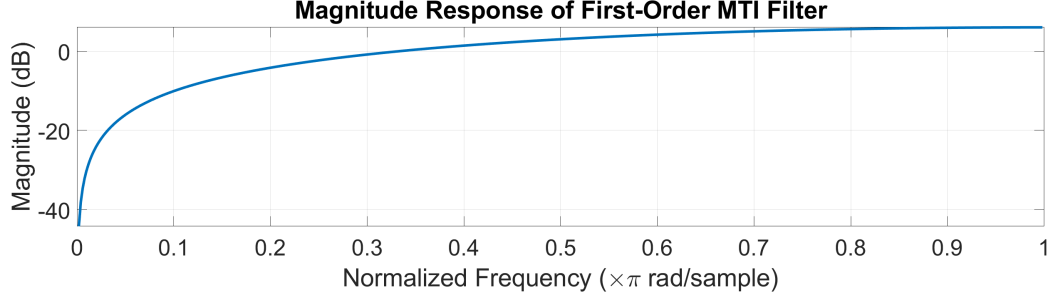


Figure 3.4: Frequency Response of First-Order MTI Filter

MTI filter may generalize to higher-order as the following

$$H_M(z) = (1 - z^{-1})^{(M-1)} \quad (3.1.5)$$

$$\hat{Y}_M[i, k - M] = \sum_{m=0}^M (-1)^m \binom{M}{m} Y_f[i, k - m].$$

The transfer function of the higher-order MTI filter is represented by $H_M(z)$, while \hat{Y}_M denotes its impulse response, where M is the filter order. However, in this study, only the first-order MTI filter is evaluated.

3.2 Spectrum Estimation

This section discusses three commonly used AOA estimation techniques. As shown in Figure 2.3, the virtual antenna array consists of $v = 12$ elements, enabling the estimation of both elevation and azimuth angles. The input to this module can be either Y obtained directly after range estimation, or \hat{Y} , which results from clutter removal.

3.2.1 Angle FFT

The phase difference between two adjacent virtual antennas for IWR6843AOP as shown in 2.3 is given as

$$\begin{aligned}\omega_z &= 2\pi \frac{\Delta d \sin \phi}{\lambda} \\ \omega_x &= 2\pi \frac{\Delta d \cos \phi \sin \theta}{\lambda},\end{aligned}\tag{3.2.1}$$

where a_x and a_z represent the phase differences along the azimuth and elevation axes. Δd denotes the spacing between adjacent virtual antennas. ϕ and θ represent the elevation and azimuth angles, respectively. "For a column of virtual antennas with a common azimuth index, the phase shift induced by the array placement is given by $\mathbf{a}_z = \{e^{j0\omega_z}, \dots, e^{j(n_z-1)\omega_z}\}$. Similarly, for a row of virtual antennas with the same elevation index, the phase shift due to the array placement is represented by $\mathbf{a}_x = \{e^{j0\omega_x}, \dots, e^{j(n_x-1)\omega_x}\}$. Two-dimensional fast Fourier transforms are applied to the virtual antenna array along the azimuth and elevation axes, respectively. After FFT, for each index k_x and k_z , the corresponding phase term can be obtained by $\omega_x = \frac{2\pi k_x}{N_{fft}}$ and $\omega_z = \frac{2\pi k_z}{N_{fft}}$, where N_{fft} is the FFT size.

Angle FFT is computationally efficient due to the use of FFT. However, Angle FFT has poor resolution with $\Delta\theta = \frac{\lambda}{N_\theta \Delta d \cos(\theta)}$, and $\Delta\phi = \frac{\lambda}{N_\phi \Delta d \cos(\phi)}$, where $\Delta\theta$ and $\Delta\phi$ denote the angular resolutions for azimuth and elevation, respectively. N_θ and N_ϕ represent the number of array elements along the azimuth and elevation axes, respectively. The FFT also suffers from spectral leakage, which can distort angle estimation.

3.2.2 MVDR

The first step of Minimum Variance Distortion-less Response (MVDR) spectrum estimation is to estimate the correlation matrix R_i for each range index i , using all chirps within one frame. Assuming V is the vector of signals across all antennas $V = \{Y_1, Y_2, \dots, Y_v\}$, define v as the number of virtual antennas.

$$R_i = \frac{1}{n_c - 1} \sum_{k=0}^{n_c} (V[i, k])(V[i, k])^H. \quad (3.2.2)$$

Another component required for MVDR estimation is the steering vector, which is derived from the geometry of the virtual antenna array [26] as shown in Figure 2.3. The steering vector represents the set of phase-delay at each virtual antenna. The first step is to define ν - μ space.

$$\begin{aligned} \mu &= \cos \phi \sin \theta \\ \nu &= \sin \phi, \end{aligned} \quad (3.2.3)$$

θ represents azimuth angle and ϕ represents elevation angle. The virtual antenna array of the IWR6843AOP radar consists of four elements along the azimuth axis and four elements along the elevation axis. The indices μ and ν are defined as follows

$$\begin{aligned} I_\mu &= [-1, -1, 0, 0, -1, -1, 0, 0, -3, -3, -2, -2] \\ I_\nu &= [-1, 0, -1, 0, -3, -2, -3, -2, -3, -2, -3, -2]. \end{aligned} \quad (3.2.4)$$

The steering vector a_s is computed as

$$a_s[n] = e^{j\pi(I_\mu[n]\mu + I_\nu[n]\nu)}, 1 \leq n \leq v. \quad (3.2.5)$$

However, as shown in Figure 2.3, RX1 and RX3 are 180° out of phase with respect to RX2 and RX4. Therefore, phase rotation is required for the corresponding virtual array elements. The phase rotation array is given by [13]

$$Ph = [-1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1]. \quad (3.2.6)$$

The phase-adjusted steering vector is represented by

$$\hat{a}_s[n] = Ph[n]a_s[n], 1 \leq n \leq v. \quad (3.2.7)$$

Finally, the spectrum is estimated by the following approach [3]

$$CS(\theta, \phi, i) = \frac{1}{\hat{a}_s^H R_i^{-1} \hat{a}_s}. \quad (3.2.8)$$

The MVDR algorithm preserves the integrity of the target signal while minimizing the contributions of noise and interference to the array output. It also provides high spatial resolution, making it effective in resolving closely spaced targets. However, MVDR is computationally intensive due to the necessity of correlation matrix inversion. Additionally, its performance is highly dependent on the accuracy of the correlation matrix estimation; any inaccuracies in this estimation can lead to degraded angle estimation performance.

3.2.3 MUSIC

Multiple Signal Classification (MUSIC) angle estimation also utilizes the correlation matrix and steering vector, similar to the MVDR method. The fundamental concept

involves performing eigen-decomposition on the correlation matrix and partitioning the eigenvectors into signal and noise subspaces. The MUSIC algorithm works by searching for steering vectors that are orthogonal to the noise subspace. However, this constraint implies that the number of signals within a single range bin cannot exceed the number of array elements.

As mentioned above the correlation matrix R_i for range bin i can be decomposed into R_i^s and R_i^e representing the correlation matrix for the signal and noise subspaces, respectively. To obtain the signal and noise subspaces, the initial step involves performing eigen decomposition to derive the eigenvalues and eigenvectors. The eigenvalues of R_i , $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_v\}$ are sorted in descending order with associated eigenvectors $E = \{e_1, e_2, \dots, e_v\}$. Assuming D represents the number of signals. Hence, the noise space is defined as $E_e = \{e_D, \dots, e_v\}$. This study applies the Minimum Description Length (MDL) test to estimate D [27].

$$L(d) = n_c(v-d) \ln \frac{\frac{1}{v-d} \sum_{k=d+1}^v \lambda_k}{(\prod_{k=d+1}^v \lambda_k)^{\frac{1}{v-d}}} \quad (3.2.9)$$

$$D = \arg \min_d \{L[d] + \frac{1}{2}(d(2v-d) + 1) \ln n_c\},$$

λ_k represents the k -th smallest eigenvalue of R_i , n_c is the number of chirps. Given the estimates of D and E_e , the spectrum is expressed as $CS(\theta, \phi, i)$ [21]

$$CS(\theta, \phi, i) = \frac{1}{\hat{a}_s^H E_e E_e^H \hat{a}_s}. \quad (3.2.10)$$

If the signal sources are correlated, the rank of R_i^s may be smaller than D , indicating that R_i^s is rank-deficient. In this case, the performance of the MUSIC algorithm

degrades. To address this issue, spatial smoothing techniques [23] are applied. Spatial smoothing involves partitioning the virtual antenna array into overlapping sub-arrays that share the same steering vector.

The 12-element virtual antenna array of the IWR6843AOP is divided into 4 sub-arrays, as shown in Figure 3.5. Each sub-array consists of 5 virtual antennas, with 3 elements aligned along the elevation axis and 3 along the azimuth axis. The spatially smoothed correlation matrix for frame i can be represented as [4, 23].

$$\hat{R}_i = \sum_{l=1}^L R_i^l. \quad (3.2.11)$$

Let L denote the number of sub-arrays, and the common steering vector used across all sub-arrays is represented by the steering vector of the first sub-array, \hat{a}_s^1 . By applying eigenvalue decomposition and (3.2.9) over \hat{a}_s^1 to acquire the noise space after smoothing \hat{E}_ϵ . The spatially smoothed coarse estimation is derived by applying (3.2.11) using the steering vector \hat{a}_s^1 and the noise subspace \hat{E}_ϵ .

In summary, the MUSIC algorithm provides high resolution and accurate spatial spectrum estimation. By isolating the noise subspace, it also ensures robust performance even when the SNR is relatively low. However, the MUSIC algorithm requires accurate estimation of sources, is computationally expensive, and is vulnerable to source correlation, which necessitates high computational load smoothing.

3.3 Target Detection

The target detection module isolates targets from the background using the spectrum CS , providing intensity information across all angles and ranges. Typically, desired

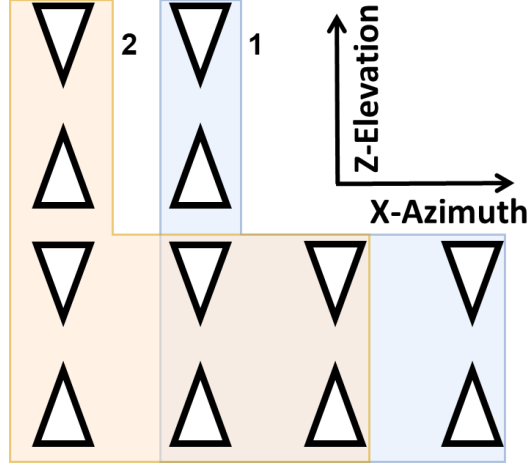


Figure 3.5: Spatial Smoothing sub-arrays

targets exhibit higher intensity values after processing. This module facilitates the extraction of points with higher intensities, effectively distinguishing them from the clutter of low-intensity points.

3.3.1 CFAR

Constant False Alarm Rate (CFAR) is a widely utilized adaptive algorithm for target detection. In this study, to circumvent the computational expense of directly applying three-dimensional CFAR to the spectrum, a **two-pass CFAR** approach is employed. This method involves a one-dimensional CFAR applied along the range dimension, followed by a two-dimensional CFAR applied to the azimuth-elevation spectrum within the corresponding range bin [12].

The CFAR algorithm iterates over each range bin for every direction (θ, ϕ) in the spatial spectrum. If the intensity of the range bin along the direction is greater than the threshold TS_r . Then the point is included in the temporary detection

list DL_1 . Then each azimuth-elevation spectrum CS_i will undergo processing by a two-dimensional CFAR. Points that exceed CFAR threshold TS_a is included in the temporary detection list DL_2 . If a point is present in both DL_1 and DL_2 , it will be appended to the output point cloud \mathcal{P} . The threshold for the CFAR algorithm is given by $TS = \alpha P_n$, where TS is the detection threshold, and α represents the scaling factor, P_n denotes the noise floor estimate. Two variations of noise floor estimation are adopted for this study

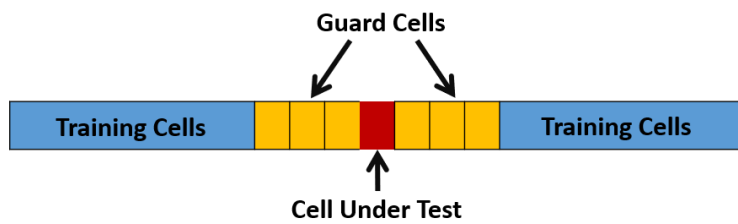
Cell averaging CFAR (CA-CFAR) typically has N_g guard cells surrounding the cell under test, as illustrated in 3.6. Guard cells are not involved in the computation of the noise floor P_n . N_t number of training cells TC is used to compute the noise floor. The dimension of noise window $N_w = N_g + N_t$.

$$P_n = \frac{1}{N_t} \sum_{m=1}^{N_t} TC_m. \quad (3.3.1)$$

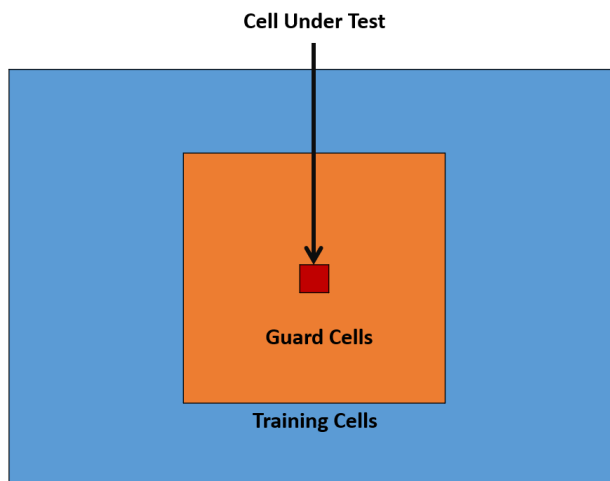
CASO-CFAR is a variation of the regular cell-averaging CFAR. Instead of averaging, it partitions the training cells into two groups, computes the average for each group, and subsequently selects the minimum.

$$P_n = \min\left(\frac{2}{N_t} \sum_{m=1}^{N_t/2} TC_m, \frac{2}{N_t} \sum_{m=N_t/2+1}^{N_t} TC_m\right), \quad (3.3.2)$$

1D-CFAR can be directly applied over range map Y . However, a more complex approach is used to apply CFAR. As illustrated in Algorithm 1, two-pass CFAR is applied over the spectrum: first, 1D CFAR is performed along the range dimension for each (θ, ϕ) direction in the spatial spectrum, followed by a second pass where 2D CFAR is applied across the Azimuth-Elevation spectrum within each range bin.



(a) 1-d CFAR.



(b) 2-d CFAR.

Figure 3.6: CUT, Guard Cells and Training Cells.

Points that appear in the results of both CFAR passes are identified as points of interest. Two-pass CFAR involves the dimensions of the noise window and guard cells for both 1D CFAR and 2D CFAR. A larger noise window dimension allows for better noise suppression, but it may also increase computational complexity.

The CFAR algorithm dynamically adjusts the detection threshold based on the intensity of cells within the noise window, enhancing its robustness in non-stationary environments. Additionally, CFAR effectively suppresses clutter in scenarios with adequate signal-to-noise ratio (SNR). However, CFAR requires meticulous tuning of

the noise window size and guard cell dimensions for different scenarios and applications; otherwise, it may fail to detect the intended targets. The computational load increases significantly as the number of dimensions grows. This is because CFAR requires the computation of the noise floor for every cell under test. As the number of dimensions increases, the number of cells under test grows exponentially with dimensionality.

Algorithm 1 Two-Pass CFAR Processing

Input CS - Range-Azimuth-Elevation data cube
Output: \mathcal{P} - Output point set
Initialize $\mathcal{P} \leftarrow \emptyset$
Initialize RPOI $\leftarrow \emptyset$
Initialize APOI $\leftarrow \emptyset$
for each (θ, ϕ) direction **do**
 RPOI $\leftarrow \text{RPOI} \cup \text{1D-CFAR}(CS(\theta, \phi))$
end for
for each range bin i **do**
 APOI $\leftarrow \text{APOI} \cup \text{2D-CFAR}(CS(i))$
end for
 $\mathcal{P} \leftarrow \text{APOI} \cap \text{RPOI}$

3.3.2 Variance-based Method

Instead of using CFAR over the spectrum to determine the output point cloud, \mathcal{P} . An alternative approach is to calculate the variance over the spectrum. If range bin i exhibits a high variance in energy distribution across the spectrum $CS(i)$, it suggests that the reflected signal's energy is spatially non-uniform, indicating the presence of a target. The variance $\mathbf{v}[i]$ is defined as

$$\mathbf{v}[i] = \frac{1}{N_\theta N_\phi - 1} \sum_{\theta} \sum_{\phi} (CS(\theta, \phi, i) - \mu_i)^2, \quad (3.3.3)$$

where N_θ and N_ϕ are number of azimuth and elevation angle bins, μ_i is the mean power across all azimuth and elevation angles for range i , $\mu_i = \frac{1}{N_\theta N_\phi} \sum_\theta \sum_\phi CS(\theta, \phi, i)$. The set of ROI (Range of Interest) is formed as

$$\text{ROI} = \{i | \mathbf{v}[i] > \xi, \forall i\}, \quad (3.3.4)$$

ξ denotes the threshold, which is the median value of the variance vector \mathbf{v} and a scaling factor are used to define the threshold, $\xi = \tau \text{Med}(\mathbf{v})$ [33].

The next step is to search the point of interest over the spatial spectrum for all range i within the set ROI. The algorithm utilizes a preset threshold TS . Additionally, when mmWave interacts with an object of interest, such as a corner reflector or human skin, it cannot penetrate the object [33].

As shown in Algorithm 2 only the closest local maximum along the range axis of direction (θ, ϕ) . Rather than directly applying 2D-CFAR exhaustively, the first local maximum with the lowest range bin is utilized, acknowledging that mmWave cannot penetrate the target of interest. This approach effectively reduces unwanted multipath noise. The predefined threshold TS governs the classification of a point as a local maximum. A lower threshold value increases the likelihood of incorporating noise points, whereas a higher threshold value risks excluding points critical for accurately representing the target.

The variance-based method exhibits reduced computational complexity and enhanced robustness to changes in scenarios. However, it is more susceptible to erroneously including artifacts that resemble local maxima.

Algorithm 2 Variance-based Method

Input: CS - Spatial spectrum for all range and angles.

TS - Preset threshold

RoI - Range of Interest (valid set of i indices)

Output: \mathcal{P} - List of detected points of interest (θ, ϕ, i)

Initialize $\mathcal{P} \leftarrow \emptyset$

for each (θ, ϕ) pair **do**

 PointFound \leftarrow False

for each i in RoI **do**

if $CS(\theta, \phi, i) > TS$ **then**

if $CS(\theta, \phi, i) > CS(\theta, \phi, i - 1)$ and $CS(\theta, \phi, i) > CS(\theta, \phi, i + 1)$ **then** \triangleright

Check for local maximum

$\mathcal{P} \leftarrow \mathcal{P} \cup CS(\theta, \phi, i)$

 PointFound \leftarrow True

break

\triangleright Exit loop since first local maximum found

end if

end if

end for

if PointFound = False **then**

 No point found for this (θ, ϕ)

end if

end for

3.3.3 Zoom-in spectrum estimation

Once \mathcal{P} is obtained, a refined zoom-in spectrum estimation can be applied around the points in \mathcal{P} , particularly if the scanning step for the angle of arrival estimation is relatively coarse. Assuming that the coarse azimuth and elevation steps are μ_{step} and ν_{step} . The step sizes for zoom-in angle estimation are $\frac{\mu_{step}}{M_{zi}}$ for azimuth $\hat{\mu}_{step}$ and $\frac{\nu_{step}}{M_{zi}}$ for elevation $\hat{\nu}_{step}$. M_{zi} is the number of zoom-in steps for each direction. For each point in the output, point set \mathcal{P} acquired from Target detection, the zoom-in steering vector derived from (3.2.4), and (3.2.5) is expressed as [12]

$$a_s[n, m_\mu, m_\nu] = e^{j\pi(I_\mu[n](\mu+m_\mu\hat{\mu}_{step})+I_\nu[n](\nu+m_\nu\hat{\nu}_{step}))}, \quad (3.3.5)$$

$$1 \leq n \leq v, -M_{zi} \leq m_\mu \leq M_{zi}, -M_{zi} \leq m_\nu \leq M_{zi},$$

ν and μ are the ν - μ space for one point p^* selected from \mathcal{P} . After obtaining the steering vectors for all antennas corresponding to the zoom-in directions, the MVDR or MUSIC algorithms can be applied to each zoom-in direction according to the approach described in the Spectrum Estimation section to compute the zoom-in spectrum \mathcal{P}_{zi} centring at p^* . Two criteria are applied to determine whether points in each zoom-in spectrum will be included in the \mathcal{P} set.

Algorithm 3 Zoom-in points detection

Input: p^* - one point in \mathcal{P} ; P_n - Noise estimate derived from noise floor of 1d
CFAR on range

$SNR_{\text{threshold}}$ - SNR threshold; T_{relative} - Relative threshold

M_{zi} - step size for zoom-in angle estimation; Ph - phase rotation array

Output: POI_{zi}

Initialize $POI_{zi} \leftarrow \emptyset$

$p_{\max} \leftarrow \max(\eta(\mathcal{P}_{zi}))$

$p_{\min} \leftarrow \min(\eta(\mathcal{P}_{zi}))$

Sharpness = $\frac{p_{\max} - p_{\min}}{p_{\max} + p_{\min}}$

for $-M_{zi} \leq m_{\mu} \leq M_{zi}$ **do**

for $-M_{zi} \leq m_{\nu} \leq M_{zi}$ **do**

$a_s = e^{j\pi(I_{\mu}(\mu+m_{\mu}\hat{\mu}_{step})+I_{\nu}(\nu+m_{\nu}\hat{\nu}_{step}))}$

$\hat{a}_s = Pha_s$

if Spectrum Estimation == MVDR **then**

$\mathbf{P}_{\text{neighbor}} = \frac{1}{\hat{a}_s^H R_i^{-1} \hat{a}_s}$

end if

if Spectrum Estimation == MUSIC **then**

$\mathbf{P}_{\text{neighbor}} = \frac{1}{\hat{a}_s^H E_{\epsilon} E_{\epsilon}^H \hat{a}_s}$

end if

if $\frac{p_{\max}}{P_n} > SNR_{\text{threshold}}$ **and** $\eta(\mathbf{p}_{\text{neighbor}}) > p_{\max} \cdot (T_{\text{relative}} - \text{Sharpness})$ **then**

 Add $POI_{zi} \leftarrow POI_{zi} \cup \mathbf{p}_{\text{neighbor}}$

end if

end for

end for

The SNR threshold determines if the entire spectrum qualifies for points of interest selection, with higher values of $T_{relative}$ leading to fewer selected points. Subsequently, Algorithm 3 is applied to each zoom-in spectrum generated from the points in \mathcal{P} . Then concatenate all POI_{zi} to \mathcal{P}

3.4 Spherical coordinates to Cartesian coordinates

After obtaining \mathcal{P} in spherical coordinates for frame f from the preceding steps, the Cartesian coordinates, with the radar as the origin, can be computed as follows:

$$\begin{aligned} x &= r \cos \phi \sin \theta \\ z &= r \sin \phi \\ y &= \sqrt{r^2 - x^2 - z^2}, \end{aligned} \tag{3.4.1}$$

x is the horizontal axis, y is the depth axis and z is the height axis. r is the range computed from the range index i , $r = i\Delta r$. If spherical coordinates \mathcal{P} are in μ - ν domain, given (3.2.3), x and z can be rewritten as

$$\begin{aligned} x &= r\mu \\ z &= r\nu. \end{aligned} \tag{3.4.2}$$

For the IWR6843AOP radar, which has an antenna spacing of $\Delta d = \frac{\lambda}{2}$ as depicted in Figure 2.3, the angular frequencies ω_x and ω_z obtained from the angle FFT can be normalized to dimensionless spatial frequency components. Based on (3.2.1), these components are redefined as $\mu = \frac{\omega_x}{\pi}$ and $\nu = \frac{\omega_z}{\pi}$. This allows computation of

Cartesian coordinates using (3.4.2). The points in \mathcal{P} are represented as $\{x, y, z, \eta(\mathbf{p})\}$ in Cartesian coordinates, where $\eta(\mathbf{p})$ is the intensity of the point \mathbf{p}

Chapter 4

Evaluation Metrics

4.1 Static Point Target Metric

For the static corner reflector scenario, Euclidean distance, range difference, and angle differences are employed as metrics. For each point cloud \mathcal{P} generated from the pipeline

$$\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_{m-1}, \mathbf{p}_m\}, \quad (4.1.1)$$

Then, all the points are sorted based on the intensity

$$\mathcal{P}_{sorted} = \{\mathbf{p}_l \in \mathcal{P} | \eta(\mathbf{p}_1) > \eta(\mathbf{p}_2) > \dots > \eta(\mathbf{p}_m)\}, \quad (4.1.2)$$

\mathcal{P}_{sorted} represents the point cloud sorted in descending order according to intensity. $\eta(\mathbf{p}_l)$ is the intensity of point \mathbf{p}_l . The subsequent step is to select a subset \mathcal{P}_h consisting of the top h number of points from \mathcal{P}_{sorted} . This method reduces the number of points involved in Euclidean distance computations while ensuring that only higher energy points, presumed to be reflections from the corner reflectors, are considered

in the error analysis.

$$\mathcal{P}_h = \{\mathbf{p}_l | 1 \leq l \leq h, \mathbf{p}_l \subseteq \mathcal{P}_{sorted}\}. \quad (4.1.3)$$

Euclidean distances $d(\mathcal{P}_h, \mathbf{p}_{gt})$ are computed for each point in \mathcal{P}_h and ground truth \mathbf{p}_{gt}

$$\mathbf{d}(\mathcal{P}_h, \mathbf{p}_{gt}) = \{d_l | d_l = \sqrt{(X_l - X_{gt})^2 + (Y_l - Y_{gt})^2 + (Z_l - Z_{gt})^2}, 1 \leq l \leq h\}, \quad (4.1.4)$$

where X_l , Y_l , and Z_l indicate the x,y, and z components of the Cartesian coordinates of \mathcal{P}_h . In the subsequent step, the index corresponding to the minimum Euclidean distance is identified by selecting the smallest value from $\mathbf{d}(\mathcal{P}_h, \mathbf{p}_{gt})$.

$$l_{min} = \arg \min_l \{d_l\}. \quad (4.1.5)$$

The point at the index of minimum Euclidean distance is defined as

$$\mathbf{p}_{lmin} = \mathcal{P}_h[l_{min}]. \quad (4.1.6)$$

Then the corresponding Euclidean distance error is $E_d = d_{lmin}$. Range, Azimuth and Elevation angle errors can also be computed via Cartesian coordinates X_{lmin} , Y_{lmin} , Z_{lmin} from \mathbf{p}_{lmin} and X_{gt} , Y_{gt} , Z_{gt} from the ground truth \mathbf{p}_{gt} .

$$\begin{aligned} \theta_{lmin} &= \arctan \frac{Y_{lmin}}{X_{lmin}} \\ \phi_{lmin} &= \arctan \frac{Z_{lmin}}{\sqrt{(X_{lmin})^2 + (Y_{lmin})^2}} \\ r_{lmin} &= \sqrt{(X_{lmin})^2 + (Y_{lmin})^2 + (Z_{lmin})^2}, \end{aligned} \quad (4.1.7)$$

$$\begin{aligned}
\theta_{gtlmin} &= \arctan \frac{Y_{gtlmin}}{X_{gtlmin}} \\
\phi_{gtlmin} &= \arctan \frac{Z_{gtlmin}}{\sqrt{(X_{gtlmin})^2 + (Y_{gtlmin})^2}} \\
r_{gtlmin} &= \sqrt{(X_{gtlmin})^2 + (Y_{gtlmin})^2 + (Z_{gtlmin})^2},
\end{aligned} \tag{4.1.8}$$

θ_{lmin} , ϕ_{lmin} , and r_{lmin} are range, azimuth angle and elevation angle for $P_{lmin,f}$. θ_{gtlmin} , ϕ_{gtlmin} , and r_{gtlmin} are range, azimuth angle and elevation angle for P_{gt} . With the numbers computed above, range error E_r , azimuth error E_θ and elevation error E_ϕ for a single frame can be calculated.

$$\begin{aligned}
E_r &= |r_{lmin} - r_{gt}| \\
E_\theta &= |\theta_{lmin} - \theta_{gt}| \\
E_\phi &= |\phi_{lmin} - \phi_{gt}|.
\end{aligned} \tag{4.1.9}$$

In the end, evaluating performance across multiple frames is required, the metric can be calculated as the average over n_t frames.

$$\overline{E}_d = \frac{1}{n_t} \sum_{f=1}^{n_t} E_{d,f}, \overline{E}_r = \frac{1}{n_t} \sum_{f=1}^{n_t} E_{r,f}, \overline{E}_\theta = \frac{1}{n_t} \sum_{f=1}^{n_t} E_{\theta,f}, \overline{E}_\phi = \frac{1}{n_t} \sum_{f=1}^{n_t} E_{\phi,f}. \tag{4.1.10}$$

If multiple corner reflectors are present, a larger number of points h from \mathcal{P}_h can be selected. Subsequently, (4.1.4) to (4.1.10) is applied iteratively to compute the errors for each corner reflector.

4.2 Human Target Metric

4.2.1 Coverage

The human target scenario introduces an additional challenge in evaluation compared to static point targets. Reflections from the human body differ from those of point targets, as they involve multiple reflection points originating from various body parts, each exhibiting similar intensity levels. These reflection points span a broad range of bins across the dimensions of range, azimuth angle, and elevation angle, resulting in a sparser point cloud representation of the human body. The point clouds vary with random body movement, respiratory activity, and pose variations. To provide information on the point cloud, an alternative approach is formulated using Euclidean distance to human mesh surface as a metric to evaluate human target scenarios.

The metric is defined as

$$\hat{D}(\mathbf{p}_l) = d(\mathbf{p}_l, \mathcal{M}), \quad (4.2.1)$$

where \mathbf{p}_l is a point within \mathcal{P} , containing only Cartesian coordinate information. $d(\mathbf{p}_l, \mathcal{M})$ is the shortest Euclidean distance from the point \mathbf{p}_l to the human mesh surface \mathcal{M} .

If $\hat{D}(\mathbf{p}_l) < 2\Delta r$, the point is classified as the reflection from the human subject, otherwise, it is considered as noise.

$$\mathcal{P}_{mesh} = \{\mathbf{p}_l \mid |\hat{D}(\mathbf{p}_l)| < 2\Delta r, \forall l\}. \quad (4.2.2)$$

To measure how well a point cloud captures reflections from a human, we can calculate the ratio of the number of points in \mathcal{P}_{mesh} to the total number of points in

the point cloud \mathcal{P} .

$$c = \frac{|\mathcal{P}_{mesh}|}{|\mathcal{P}|}. \quad (4.2.3)$$

Consider a set of joints $\mathcal{J} = \{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_\kappa\}$, where κ is the total number of joints. For $p \in \mathcal{P}_{mesh}$, find the closest joint by the minimum Euclidean distance.

$$\text{closest_joint}(p) = \arg \min_{\mathbf{j} \in \mathcal{J}} \|p - \mathbf{j}\|. \quad (4.2.4)$$

Define subsets $\mathcal{P}_l \subseteq \mathcal{P}_{mesh}$ where \mathcal{P}_l is the set of points closest to joint \mathbf{j}_l :

$$\mathcal{P}_l = \{p \in \mathcal{P}_{mesh} : \text{closest_joint}(p) = j_l\}. \quad (4.2.5)$$

Define a mapping $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_s\}$, where each element corresponds to a body part. The set includes five body parts: the upper torso, lower torso, head, legs, and arms. By grouping the joints into the body parts subset based on the anatomical position of the joints, the coverage of each body segment is computed as the proportion of points within each segment to the total number of points in \mathcal{P}_{mesh} .

$$c_k = \frac{\sum_{l \in b_\rho} |\mathcal{P}_l|}{|\mathcal{P}_{mesh}|}, 1 \leq \rho \leq s. \quad (4.2.6)$$

4.2.2 Consistency

Another critical metric for evaluating human target point cloud generation is assessing its consistency. For the same scenario, if the subject maintains a consistent location in Cartesian space, a stable orientation relative to the radar origin and the same body pose in consecutive frames, such as when sitting in a chair without moving it or changing direction, then the corresponding point clouds for any two frames should

exhibit similar intensities and distributions. Two metrics used to assess intensities are the maximum and mean values for each point cloud. This grouping facilitates the identification of body parts with higher point densities in the point cloud, offering insights into the performance of point cloud generation pipelines.

$$I_{cloud} = \eta(\mathcal{P})$$

$$I_{max} = \max(I_{cloud}) \tag{4.2.7}$$

$$I_{mean} = \frac{1}{m} \sum_{l=1}^m I_l.$$

$$\sigma_{max} = \frac{\sqrt{I_{f,max} - \bar{I}_{max}}}{n_f}$$

$$\sigma_{mean} = \frac{\sqrt{I_{f,mean} - \bar{I}_{mean}}}{n_f}. \tag{4.2.8}$$

To assess consistency in point cloud distribution, **Hausdorff distance** d_H between the two point clouds is utilized. For two point clouds \mathcal{P}_f at frame index f , and \mathcal{P}_{f+j} at frame index $f + j$. The Hausdorff distance $d_H(\mathcal{P}_f, \mathcal{P}_{f+j})$ can be defined as the following [25].

$$A = \mathcal{P}_f$$

$$B = \mathcal{P}_{f+j} \tag{4.2.9}$$

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} \| \mathbf{a} - \mathbf{b} \|_2, \sup_{b \in B} \inf_{a \in A} \| \mathbf{a} - \mathbf{b} \|_2 \right\}.$$

A smaller $d_H(\mathcal{P}_f, \mathcal{P}_{f+j})$ indicates greater consistency between the two point clouds from different frames.

Chapter 5

Performance Evaluation

This chapter outlines the evaluation scenarios and their corresponding metrics. Section 5.1 details the testbed setup, while Section 5.2 explains ground truth generation for both static point target and human target scenarios. Experimental trials are conducted to assess the performance of various method combinations and to demonstrate how the proposed metrics provide insights into the quality of the point cloud generated by different processing pipelines. Section 5.3 presents the results of the static target test, and Section 5.4 concludes with the results from the human subject test.

5.1 Testbed Setup

5.1.1 Radar

A TI IWR6843AOP FMCW radar and a DCA1000EVM [11] are used in the experiment. As shown in Table 2.1, the radar configuration and specifications remain consistent across all frames and trials. While the TI IWR6843AOP provides a relatively

high range resolution due to its large bandwidth, the 4.41 cm range resolution may still introduce errors when compared to the ground truth. Additionally, the limited number of virtual antennas results in a 29-degree angular resolution in both azimuth and elevation, which can increase Euclidean distance errors at greater ranges. The reduced angular resolution also limits the radar’s ability to resolve multiple targets at the same range. Furthermore, the radar is susceptible to multi-path effects. Since the lab room is not fully covered with foam absorbers, artifacts may arise due to multipath reflections. These signals can interfere with the direct-path signal, altering its phase and leading to biased spectrum estimation.

5.1.2 Corner Reflector

For static point target scenarios, one or multiple corner reflectors are placed within the FOV of the radar. Ground truth locations of corner reflectors are collected from the motion capture system - OptiTrack motive motion capture system [17] by twelve cameras with markers attached to corner reflectors and the radar.

The experiment was conducted in an indoor laboratory environment measuring 6.5 m by 6 m. However, the placement of infrared cameras and the radar was configured to ensure the cameras could accurately capture the radar’s position. The radar experiment zone is 5.5 m by 5 m. Foam absorbers are employed to mitigate reflections from the floor surface in proximity to the placement of the corner reflector. The corner reflector is a polyhedron with six hollow equilateral triangles. As demonstrated in Figure 5.1, each edge has a length of 21 cm. Each hollow equilateral triangle consists of three solid isosceles triangles, where each has two equal edges measuring 14.8 cm in length.

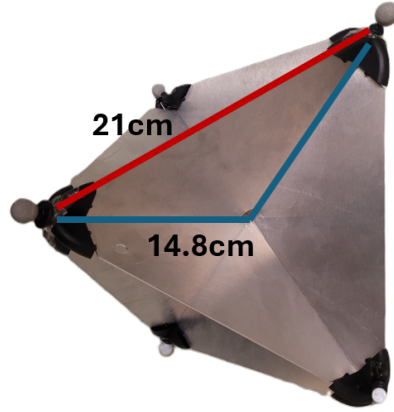


Figure 5.1: Corner reflector marker placement

For single corner reflector, experiments are conducted in close range ($r < 1\text{m}$), medium range ($1\text{m} < r < 3\text{m}$), and far range ($r > 3\text{m}$). Figure 5.4 shows examples of close, medium and far-range placements of the corner reflector.

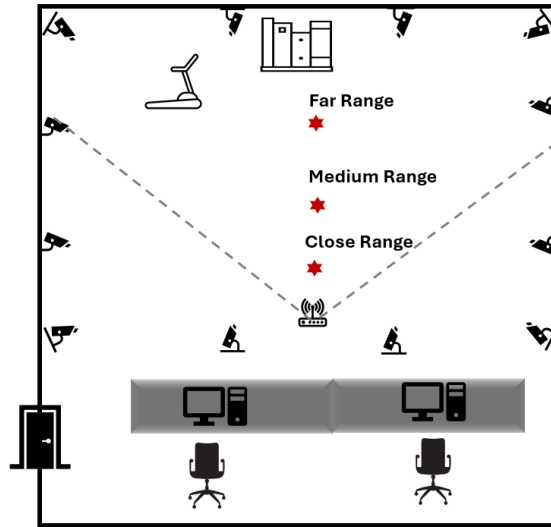


Figure 5.2: Lab space setup for single corner reflector

In the scenario involving three corner reflectors, the IWR6843AOP radar was positioned parallel to the floor, oriented downward. The radar antennas face the

floor, and three corner reflectors are placed on the floor as shown in Figure 5.3. Each corner reflector is positioned at a minimum distance of 30 cm from the centre of any adjacent corner reflector. A 30 cm distance from the center is maintained to prevent physical overlap between the corner reflectors.

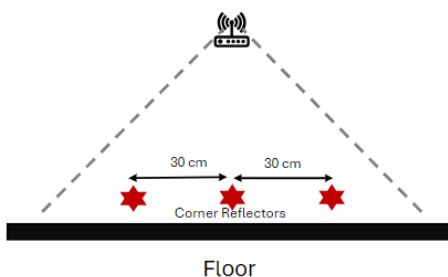


Figure 5.3: Three corner reflectors scenario setup

5.1.3 Human Subject

The same radar setup is used in the experiment. A motion capture system is also employed to obtain the test subject's ground truth positions and joint angles. Markers are attached to a motion capture suit worn by the test subject. The experiment is conducted in the same laboratory environment as the corner reflector scenario, involving a subject with a height of 183 cm. The movement of the subject is constrained within an area defined by the radar's field of view (FOV). For simplicity, this area is represented as a green rectangle in the lab space, as illustrated in Figure 5.4.

Five scenarios are included for the human subject experiment: Standing, Walking in place, Walking, Dancing, and Sitting.

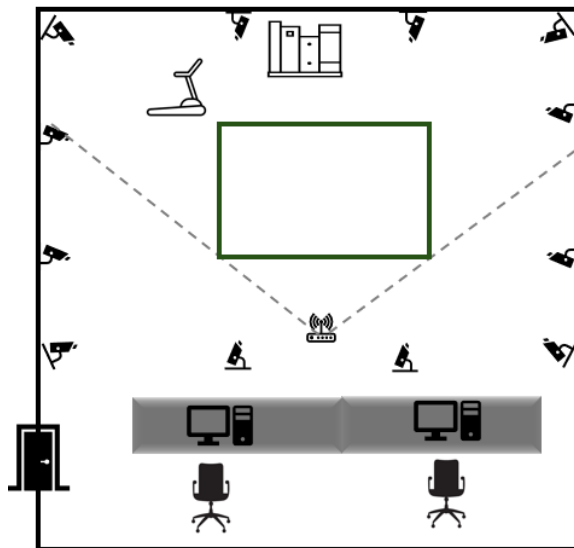


Figure 5.4: Human subject scenario setup

5.2 Ground Truth Generation

As mentioned in the previous section, ground truth is acquired by using the motion capture system. As shown in Figure 5.2 and Figure 5.4, there are 12 motion capture cameras. After system calibration, the first step for both corner reflector scenario and human subject scenario is to determine the position and rotations of radar. Markers and radar are placed as follow

As demonstrated in Figure 5.5, four markers indicated by green hollow circles are placed parallel to the radar board. Four markers constitute a rigid body within the motion capture system, providing both position and rotation data for the radar. The corner reflector has markers attached to it forming a rigid body as shown in Figure 5.6.

The rigid body of the corner reflector defines its ground truth global coordinate position, \hat{P}_{gt} . The radar's global coordinate, \hat{P}_{radar} , and rotation matrix, R_{radar} ,

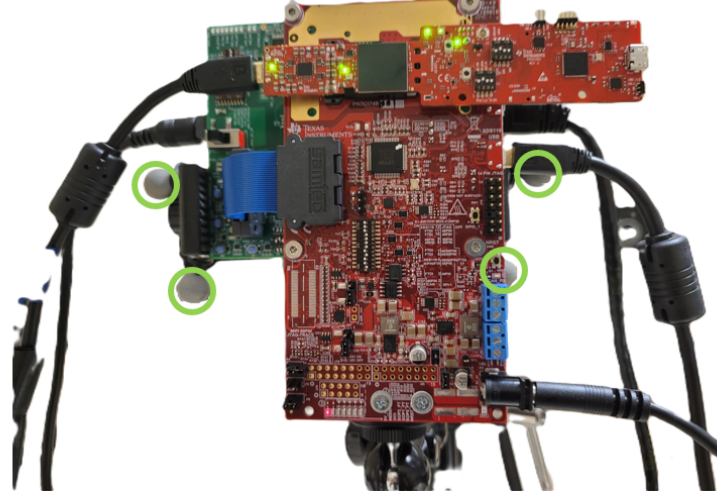


Figure 5.5: Radar and marker placement

are derived from its rigid body. To compute the ground truth position of the corner reflector in the local coordinate frame, the first step involves calculating its relative position in global coordinates. Then, the ground truth location, P_{gt} , in the radar's local coordinate frame is obtained by applying the appropriate transformation.

$$\begin{aligned}\hat{P}_{rel} &= \hat{P}_{gt} - \hat{P}_{\text{radar}} \\ P_{gt} &= R^{-1} \hat{P}_{rel}\end{aligned}\tag{5.2.1}$$

In the human target scenario, the test subject wears a motion capture suit equipped with markers placed at joint and bone locations as shown in Figure 5.7. The motion capture system tracks the markers to generate a skeletal model that accurately reflects the subject's movements. subsequently, the skeleton model from motion capture system is re-targeted to Skinned Multi-Person Linear model (SMPL) skeleton model to obtain the pose parameter $\vec{\theta}$. SMPL is a skinned, vertex-based model designed to represent a wide range of body shapes in natural human poses with an

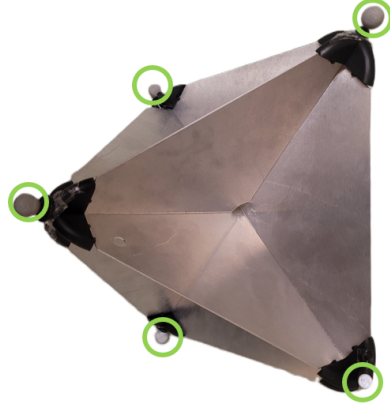


Figure 5.6: Corner reflector marker placement

underlying skeletal structure [15]. The global translation \vec{P}_T is determined by the position of the pelvis joint relative to the radar's local coordinate frame, as calculated in (5.2.1). With pose independent shape parameter $\vec{\beta}$, SMPL human mesh $\mathcal{M}(\vec{\theta}, \vec{\beta}, \vec{P}_T)$ model is generated [15].



(a) Front



(b) Back

Figure 5.7: Motion capture suit and marker placement

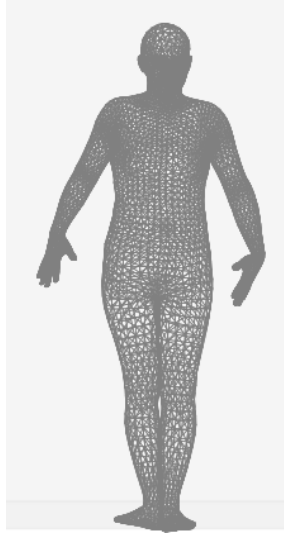


Figure 5.8: An Example of SMPL Human Mesh

5.3 Static Point Target Test

Clutter removal is not applied, as the corner reflector is a static target. As illustrated in Figure 3.2, the Angle FFT, MVDR, and MUSIC algorithms are evaluated for the angle estimation module. Additionally, both the variance-based method and the CFAR method are assessed.

Method	Spectrum Estimation	Target Detection
Method 1	MVDR	Two-pass CFAR(CASO) + Zoom-in spectrum estimation

Continued on the next page

Continued from previous page

Method	Spectrum Estimation	Target Detection
Method 2	MVDR	Two-pass CFAR(CA) + Zoom-in spectrum estimation
Method 3	MVDR	Variance-Based method
Method 4	MUSIC	Variance-Based method + 1D Range CFAR
Method 5	Angle FFT	Variance-Based method

Table 5.1: Pipelines for Static Target test

CFAR used for static target test is defined in Table 5.2

1D-CFAR Dimension of Noise Window	8
1D-CFAR Dimension of Guard Cells	4
2D-CFAR Dimension of Noise Window	8
2D-CFAR Dimension of Guard Cells	4

Table 5.2: CFAR settings for Static Target Test

5.3.1 Single-Target Test

As shown in Table 5.3, ten scenarios are included for the single-target test, ground truth location relative to the radar obtained from Motion Capture in Cartesian coordinates.

Trial	X	Y	Z
1	-0.019 m	0.538 m	0.040 m
2	-0.018 m	1.305 m	0.062 m
3	-0.391 m	1.302 m	0.066 m
4	-0.3 m	0.464 m	0.021 m
5	-0.037 m	0.941 m	0.161 m
6	-0.373 m	0.910 m	0.158 m
7	-0.438 m	1.415 m	0.149 m
8	0.173 m	2.496 m	-0.074 m
9	-0.63 m	2.535 m	-0.095 m
10	0.266 m	3.632 m	-0.198 m

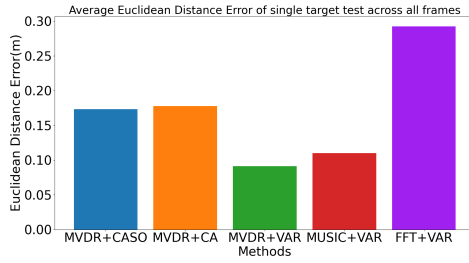
Table 5.3: Ground Truth of Trials for Single-target Test

Figure 5.9 illustrates average errors across all trials. Figure 5.10 to Figure 5.13 visualize Euclidean distance, range, azimuth and elevation errors of all trials and methods. The bar represents the mean error and error bar represent the minimum and maximum error for respective trial and method. Table 5.4 to Table 5.13 illustrate mean Euclidean distance error \overline{E}_d , range error \overline{E}_r , azimuth angle error \overline{E}_θ , elevation angle error \overline{E}_ϕ , and their corresponding standard deviations σ_d , σ_r , σ_θ , σ_ϕ .

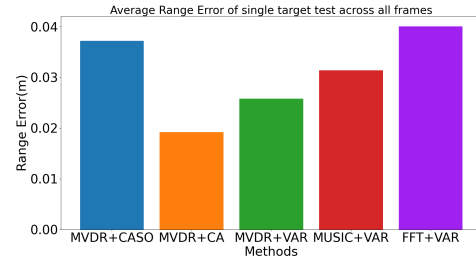
The evaluation results indicate that all methods demonstrate improved performance when the corner reflector is positioned at a closer distance and when the azimuth and elevation angles relative to the radar are smaller. Signal intensity experiences attenuation as the range increases, following the path loss model given by $L_r = 20 \log(4\pi r/\lambda)$ [7]. The attenuation of signal strength at greater ranges reduces the ability of pipelines to distinguish the target from the background, thereby contributing to the degradation of the pipelines' performance. The accuracy of spectrum estimation deteriorates due to the non-linear behavior of the phase term associated with the angle of arrival, as described in (3.2.1). The sensitivity of the phase term ω_z to changes in the elevation angle ϕ decreases as ϕ increases. Similarly, for a fixed elevation angle ϕ , the phase term ω_x also exhibits reduced sensitivity to changes in ϕ as ϕ becomes larger.

The Angle FFT in FFT+Var yields the worst Euclidean distance error compared to MVDR+Var and MUSIC+Var, both of which employ the variance-based method for target detection. This is attributed to the larger angle estimation errors associated with Angle FFT relative to MVDR and MUSIC as shown in Figure 5.12 and Figure 5.13. Angle FFT directly applies FFT across virtual antenna channels, leading to higher sidelobes and inherent limitations imposed by the configuration of the virtual antenna array, resulting in inaccurate spectrum estimation. In contrast, MVDR adaptively optimizes beamforming weights to suppress interference and minimize power from undesired directions, while MUSIC leverages the noise subspace to enhance resolution and eliminate noise effects. MVDR+Var and MUSIC+Var have similar Euclidean distance errors.

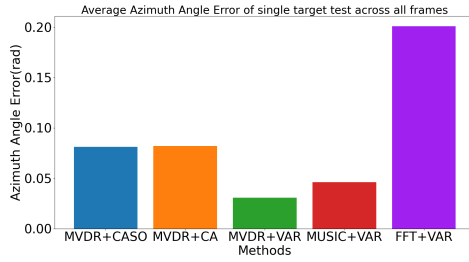
The MVDR+CASO and MVDR+CA methods compare CASO-CFAR with CA-CFAR, with MVDR+CASO demonstrating overall better performance than MVDR+CA. However, in certain instances, CASO-CFAR results in higher range estimation errors, as illustrated in Figure 5.10. Overall, the variance-based method demonstrates superior performance compared to the two-pass CFAR methods. Specifically, the MVDR+Var method, which utilizes the variance-based approach, outperforms MVDR+CASO and MVDR+CA in Trials 3, 4, 6, 7, and 9, where the azimuth or elevation angles are larger. This is attributed to the antenna gain pattern of the TI IWR6843AOP radar [13], which has a higher gain at the foresight and a lower gain at larger incident angles. For a CFAR algorithm employing dynamic thresholding, variations in antenna gain can introduce inconsistencies in threshold adaptation, potentially reducing the algorithm's effectiveness in suppressing noise and background clutter, particularly at boresight where the gain is highest. This effect becomes more pronounced when detecting targets at larger incident angles, where the lower antenna gain further weakens the received signal, leading to degraded detection performance and non-uniform false alarm rates.



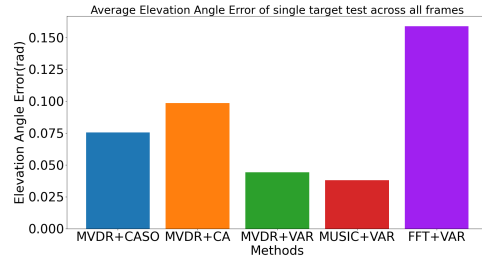
(a) Euclidean Distance Error



(b) Range Error



(c) Azimuth Angle Error



(d) Azimuth Angle Error

Figure 5.9: Bar Graph of Average Error across All Trials

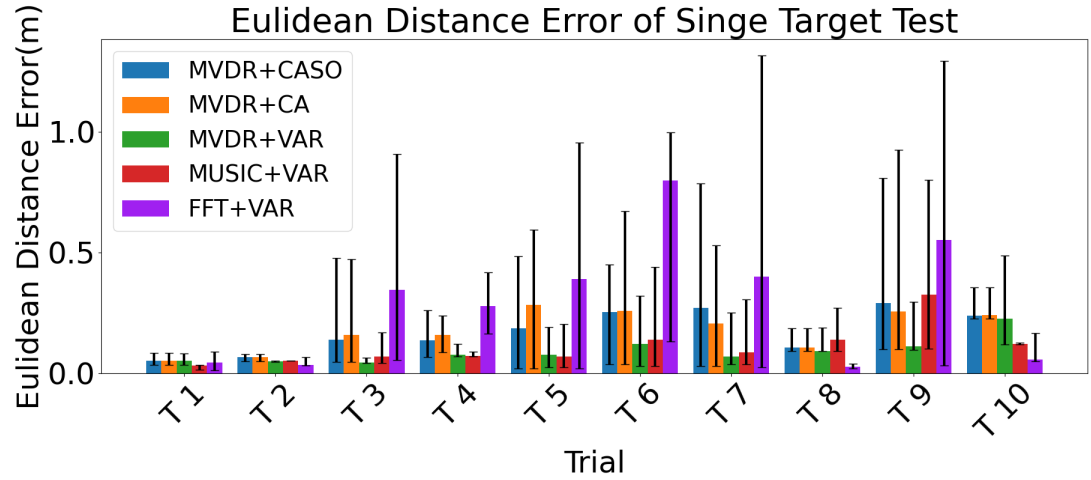


Figure 5.10: Euclidean Distance Error of Single-target Test.

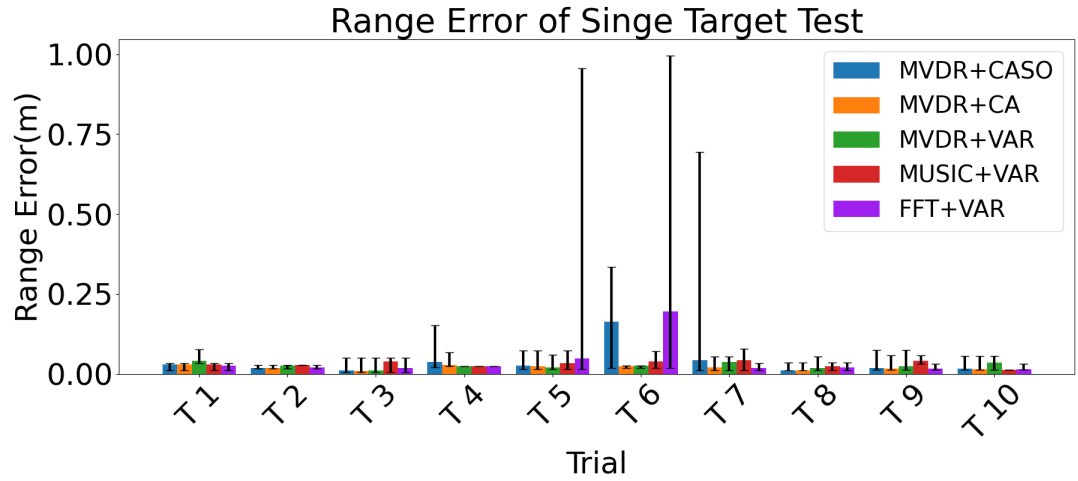


Figure 5.11: Range Error of Single-target Test.

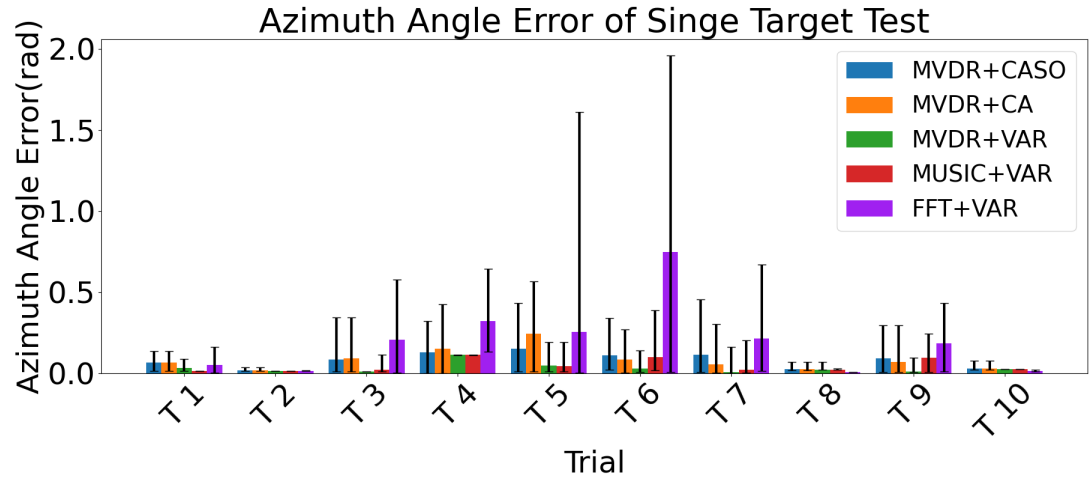


Figure 5.12: Azimuth Angle Error of Single-target Test.

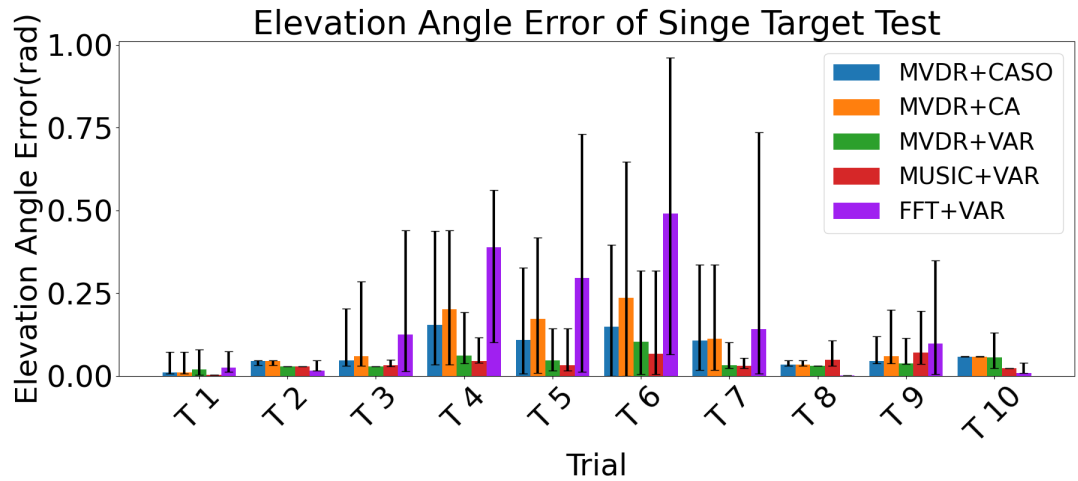


Figure 5.13: Elevation Angle Error of Single-target Test.

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	5.10 cm	5.10 cm	5.11 cm	3.14 cm	4.47 cm
σ_d	1.53 cm	1.53 cm	1.57 cm	0.72 cm	1.78 cm
\overline{E}_r	2.97 cm	2.97 cm	4.05 cm	3.03 cm	2.52 cm
σ_r	0.82 cm	0.82 cm	1.63 cm	0.77 cm	1.08 cm
\overline{E}_θ	0.066 rad	0.066 rad	0.032 rad	0.014 rad	0.052 rad
σ_θ	0.041 rad	0.041 rad	0.026 rad	0 rad	0.041 rad
\overline{E}_ϕ	0.01 rad	0.01 rad	0.019 rad	0.002 rad	0.024 rad
σ_ϕ	0.015 rad	0.015 rad	0.032 rad	0 rad	0.019 rad

Table 5.4: Mean and Standard Deviation of Trial 1 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	6.54 cm	6.54 cm	4.9 cm	5.03 cm	3.49 cm
σ_d	0.4 cm	0.41 cm	0.22 cm	0 cm	0.45 cm
\overline{E}_r	1.73 cm	1.73 cm	2.49 cm	2.8 cm	2.03 cm

Continued on the next page

Continued from previous page

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
σ_r	0.37 cm	0.37 cm	0.53 cm	0 cm	0.58 cm
\overline{E}_θ	0.016 rad	0.016 rad	0.014 rad	0.014 rad	0.015 rad
σ_θ	0.004 rad	0.004 rad	0 rad	0 rad	0.002 rad
\overline{E}_ϕ	0.045 rad	0.045 rad	0.029 rad	0.0029 rad	0.015 rad
σ_ϕ	0.002 rad	0.002 rad	0 rad	0 rad	0.003 rad

Table 5.5: Mean and Standard Deviation of Trial 2 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	13.78 cm	15.83 cm	4.29 cm	6.89 cm	34.54 cm
σ_d	10.16 cm	12.38 cm	0.65 cm	2.27 cm	23.38 cm
\overline{E}_r	1.09 cm	0.93 cm	0.99 cm	3.94 cm	1.86 cm
σ_r	1.26 cm	1.11 cm	1.24 cm	1.08 cm	1.86 cm
\overline{E}_θ	0.084 rad	0.091 rad	0.009 rad	0.019 rad	0.204 rad

Continued on the next page

Continued from previous page

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
σ_θ	0.074 rad	0.084 rad	0 rad	0.022 rad	0.161 rad
\overline{E}_ϕ	0.047 rad	0.059 rad	0.028 rad	0.032 rad	0.015 rad
σ_ϕ	0.032 rad	0.054 rad	0 rad	0.008 rad	0.124 rad

Table 5.6: Mean and Standard Deviation of Trial 3 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	13.54 cm	15.72 cm	7.52 cm	7.03 cm	27.75 cm
σ_d	4.95 cm	4.7 cm	1.16 cm	0.51 cm	6.79 cm
\overline{E}_r	3.65 cm	2.72 cm	2.39 cm	2.39 cm	2.39 cm
σ_r	3.24 cm	1.16 cm	0 cm	0 cm	0 cm
\overline{E}_θ	0.128 rad	0.148 rad	0.113 rad	0.113 rad	0.32 rad
σ_θ	0.093 rad	0.1 rad	0.001 rad	0.001 rad	0.131 rad
\overline{E}_ϕ	0.154 rad	0.201 rad	0.06 rad	0.044 rad	0.389 rad

Continued on the next page

Continued from previous page

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
σ_ϕ	0.116 rad	0.136 rad	0.038 rad	0.02 rad	0.14 rad

Table 5.7: Mean and Standard Deviation of Trial 4 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	18.67 cm	28.26 cm	7.56 cm	6.97 cm	39.04 cm
σ_d	15.13 cm	19.83 cm	5.69 cm	5.39 cm	25.29 cm
\overline{E}_r	2.61 cm	2.05 cm	3.31 cm	2.39 cm	2.39 cm
σ_r	1.76 cm	1.28 cm	0.96 cm	1.94 cm	15.94 cm
\overline{E}_θ	0.151 rad	0.243 rad	0.048 rad	0.044 rad	0.256 rad
σ_θ	0.159 rad	0.203 rad	0.055 rad	0.054 rad	0.325 rad
\overline{E}_ϕ	0.109 rad	0.172 rad	0.047 rad	0.033 rad	0.296 rad
σ_ϕ	0.084 rad	0.123 rad	0.042 rad	0.031 rad	0.189 rad

Table 5.8: Mean and Standard Deviation of Trial 5 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	25.3 cm	25.66 cm	12.12 cm	13.94 cm	79.65 cm
σ_d	12.78 cm	18.88 cm	11.7 cm	11.11 cm	20.7 cm
\overline{E}_r	16.39 cm	2.29 cm	2.51 cm	3.97 cm	19.45 cm
σ_r	12.86 cm	0.4 cm	0.27 cm	2.27 cm	37.44 cm
\overline{E}_θ	0.111 rad	0.082 rad	0.028 rad	0.099 rad	0.749 rad
σ_θ	0.093 rad	0.076 rad	0.024 rad	0.099 rad	0.612 rad
\overline{E}_ϕ	0.148 rad	0.236 rad	0.102 rad	0.067 rad	0.491 rad
σ_ϕ	0.076 rad	0.189 rad	0.13 rad	0.08 rad	0.256 rad

Table 5.9: Mean and Standard Deviation of Trial 6 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	25.1 cm	20.62 cm	6.82 cm	8.53 cm	39.98 cm
σ_d	17.01 cm	13.77 cm	3.62 cm	6.06 cm	34 cm
\overline{E}_r	4.18 cm	1.93 cm	3.65 cm	4.2 cm	1.89 cm

Continued on the next page

Continued from previous page

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
σ_r	9.6 cm	1.55 cm	2.15 cm	2.36 cm	1.14 cm
\overline{E}_θ	0.115 rad	0.053 rad	0.007 rad	0.02 rad	0.212 rad
σ_θ	0.109 rad	0.069 rad	0.022 rad	0.05 rad	0.157 rad
\overline{E}_ϕ	0.107 rad	0.112 rad	0.032 rad	0.031 rad	0.141 rad
σ_ϕ	0.081 rad	0.086 rad	0.015 rad	0.013 rad	0.208 rad

Table 5.10: Mean and Standard Deviation of Trial 7 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	10.49 cm	10.49 cm	9.23 cm	13.84 cm	2.75 cm
σ_d	2.98 cm	2.98 cm	0.76 cm	7.73 cm	0.91 cm
\overline{E}_r	1.06 cm	1.06 cm	1.8 cm	2.45 cm	2 cm
σ_r	0.41 cm	0.41 cm	1.25 cm	1.19 cm	1.2 cm
\overline{E}_θ	0.023 rad	0.023 rad	0.02 rad	0.02 rad	0.007 rad

Continued on the next page

Continued from previous page

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
σ_θ	0.016 rad	0.016 rad	0.003 rad	0.002 rad	0 rad
\overline{E}_ϕ	0.033 rad	0.033 rad	0.03 rad	0.049 rad	0.002 rad
σ_ϕ	0.002 rad	0.002 rad	0 rad	0.033 rad	0 rad

Table 5.11: Mean and Standard Deviation of Trial 8 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	28.99 cm	25.63 cm	11.19 cm	32.54 cm	55.23 cm
σ_d	21.17 cm	22.34 cm	3.66 cm	24.43 cm	36.7 cm
\overline{E}_r	1.83 cm	1.72 cm	2.33 cm	4.1 cm	1.69 cm
σ_r	0.41 cm	0.9 cm	1.12 cm	1.23 cm	0.78 cm
\overline{E}_θ	0.091 rad	0.068 rad	0.01 rad	0.094 rad	0.183 rad
σ_θ	0.02 rad	0.083 rad	0.016 rad	0.088 rad	0.127 rad
\overline{E}_ϕ	0.044 rad	0.06 rad	0.038 rad	0.07 rad	0.098 rad

Continued on the next page

Continued from previous page

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
σ_ϕ	0.02 rad	0.0044 rad	0.011 rad	0.054 rad	0.081 rad

Table 5.12: Mean and Standard Deviation of Trial 9 for all methods in single-target static test

	MVDR + CASO	MVDR + CA	MVDR + Var	MUSIC + Var	FFT + Var
\overline{E}_d	23.88 cm	23.94 cm	22.57 cm	11.97 cm	5.61 cm
σ_d	2.68 cm	2.66 cm	5.74 cm	0.11 cm	1.52 cm
\overline{E}_r	1.66 cm	1.51 cm	3.55 cm	1.21 cm	1.43 cm
σ_r	1.32 cm	1.09 cm	1.79 cm	0 cm	0.62 cm
\overline{E}_θ	0.028 rad	0.028 rad	0.024 rad	0.024 rad	0.012 rad
σ_θ	0.012 rad	0.012 rad	0.0006 rad	0.0004 rad	0.004 rad
\overline{E}_ϕ	0.058 rad	0.058 rad	0.055 rad	0.022 rad	0.008 rad
σ_ϕ	0.0004 rad	0.0004 rad	0.017 rad	0 rad	0.003 rad

Table 5.13: Mean and Standard Deviation of Trial 10 for all methods in single-target static test

5.3.2 Three-Target Test

The three-target test is conducted with three corner reflectors positioned as shown in Figure 5.3. Table 5.14 shows the ground truth positions of the corner reflectors relative to the radar. Corner Reflector 1 is positioned in the middle, while Corner Reflectors 2 and 3 are on the sides.

Corner Reflector	X	Y	Z
1	-0.048 m	0.79 m	0.123 m
2	-0.357 m	0.80 m	0.12 m
3	0.22 m	0.81 m	0.19 m

Table 5.14: Ground Truth of Trials for Three-target Test

As shown in Figure 5.14, Corner Reflector 1 exhibits a lower Euclidean distance error, attributed to its significantly smaller azimuth angle. This observation reinforces the conclusion from the single-target test, confirming that angle estimation accuracy deteriorates as the azimuth angle θ increases. MVDR+CASO and MVDR+CA, which utilize the two-pass CFAR approach, exhibit poorer performance than MVDR+Var across all corner reflectors. In particular, the errors are significantly higher for Corner Reflector 2 and Corner Reflector 3.

For MVDR+Var, MUSIC+Var, and FFT+Var, which use the variance-based

method for target detection but differ in their spectrum estimation algorithms. MUSIC+Var shows a significant performance degradation during the three-target test compared to the single-target test. CR2 and CR3 are positioned on the sides, centered at adjacent range bins. However, the physical size of the corner reflectors and FFT leakage during the Range FFT introduce inaccuracies in estimating the number of signals for MUSIC, leading to performance degradation. The pipelines employing CFAR produce inferior results compared to those using the variance-based algorithm, further reinforcing the conclusions drawn from the single-target test.

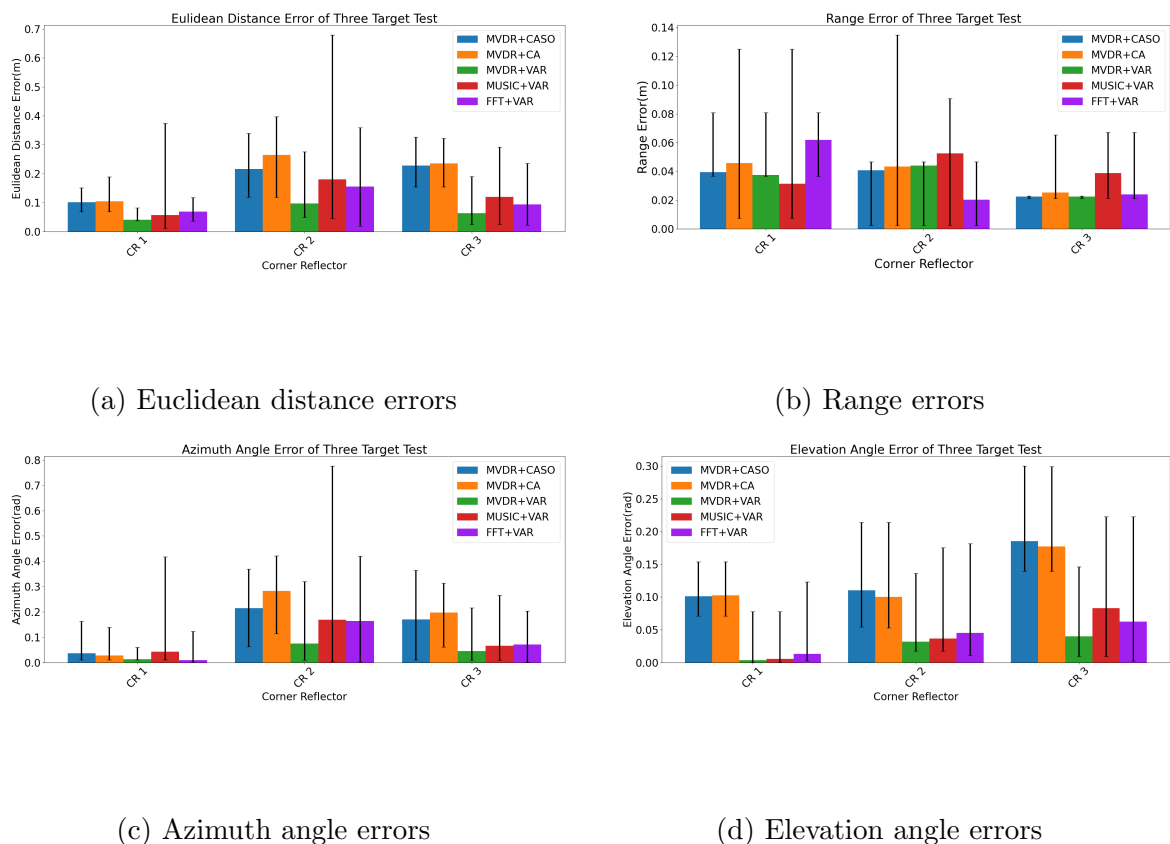


Figure 5.14: Bar Plot of Mean Errors Across Corner Reflectors and All Methods in the Three-Target Test Scenario

5.4 Human Subject Test

Four different pipelines are considered for the human subject test. All of them used MVDR spectrum estimation method. Angle FFT exhibited erroneous angle estimation when the target presented a large incident angle relative to the radar, as illustrated in Figure 5.12 and Figure 5.13. For MUSIC algorithm, it is required to know the number of incoming signals D . D is needed to be smaller than the number of antennas in the smoothed virtual array. However, as shown in Figure 3.5, the number of antennas is 5 and the value of D could exceed the number of antennas. Hence, only MVDR is considered for the human subject test.

Method	Spectrum Estimation	Clutter Removal	Target Detection
Method 1	MVDR	Mean Removal	CFAR(CASO) + Zoom-in spectrum estimation
Method 2	MVDR	MTI	CFAR(CASO) + Zoom-in spectrum estimation
Method 3	MVDR	MTI	Variance-Based method
Method 4	MVDR	Mean Removal	Variance-Based method

Continued on the next page

Continued from previous page

Method	Spectrum Estimation	Clutter Removal	Target Detection
---------------	--------------------------------	------------------------	-------------------------

Table 5.15: Pipelines for Human Subject test

5.4.1 Coverage test

To determine the optimal parameters for human subject detection, various parameter combinations were tested using the Mean+CASO method in the standing scenario.

Parameter Settings	1D-CFAR Dimension of Noise Window	1D-CFAR Dimension of Guard Cells	2D-CFAR Dimension of Noise Window	2D-CFAR Dimension of Guard Cells
Setting 1	5	2	5	2
Setting 2	5	2	8	4
Setting 3	8	4	8	4
Setting 4	8	4	14	8
Setting 5	14	8	14	8
Setting 6	16	10	16	10

Continued on the next page

Continued from previous page

Parameter	1D-CFAR	1D-CFAR	2D-CFAR	2D-CFAR
	Dimension	Dimension	Dimension	Dimension
	of Noise	of Guard	of Noise	of Guard
	Window	Cells	Window	Cells

Table 5.16: Different CFAR parameter settings

Table 5.17 presents the proportion of points identified as reflections from the human body for all CFAR configurations listed in Table 5.16. Setting 3, which was used in the single-point target test, exhibits relatively lower performance with human subjects. As shown in the table, configurations with larger noise windows generally yield better performance. Setting 5 achieves the highest proportion that satisfies the condition $\hat{D}(P_i) < 2\Delta r$, where $\Delta r = 4.41\text{cm}$ and will be used for subsequent evaluations.

Scenario	Setting	Setting	Setting	Setting	Setting	Setting
	1	2	3	4	5	6
Standing	0%	11.75%	42.36%	64.02%	74.03%	73.08%

Table 5.17: the proportion of points identified as reflections from human body for different parameter settings

Scenario	Setting	Setting	Setting	Setting	Setting
	2	3	4	5	6
Point	0.38	0.15	0.284	0.199	0.177
Cloud					
\mathcal{P}_{mesh}	1.87	1.78	0.573	0.369	0.369

Table 5.18: Hausdorff of points identified as reflections from human body for different parameter settings

Table 5.19 reports the percentage of points that satisfy the condition $\hat{D}(P_i) < 2\Delta r$, where $\Delta r = 4.41\text{cm}$ representing the proportion of points identified as reflections from the human subject. Table 5.20 reports the average of total number of points in the point cloud across all frames for each pipeline.

Scenario	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Standing	74.03%	67.35%	75.22%	69.44%
Walking in place	65.69%	70.67%	65.92%	58.46%
Walking	73.85%	77.03%	75.72%	74.02%
Dancing	73.96%	65.01%	70.71%	68.04%
Sitting	71.81%	63.20%	61.64%	62.93%

Continued on the next page

Continued from previous page

Scenario	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
-----------------	--------------------	-------------------	------------------	-------------------

Table 5.19: proportion of points identified as reflections from human body

Scenario	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Standing	297	279	43	41
Walking in place	242	217	39	45
Walking	210	197	29	34
Dancing	223	200	35	53
Sitting	267	236	42	42

Table 5.20: Average number of points in point clouds across all frames

The proportion of points identified as reflections serves as a valuable indicator of how effectively each pipeline removes noise and reduces artifacts across different scenario setups. As shown in Table 5.19, Mean+CASO and MTI+Var exhibit superior performance compared to MTI+CASO and Mean+Var in most conditions. Mean+CASO demonstrates the highest performance in scenarios where the human subject remains relatively stationary, as evidenced in the Sitting Scenario. The first-order MTI filter operates by subtracting the signal at the previous chirp index from

the current one, which may suppress slowly moving targets when variations within a frame are minimal. In contrast, the mean removal algorithm treats the mean of samples across the slow-time axis as background clutter, better preserving subtle variations. The variance-based method is prone to including noise that appears as the closest local maximum in the point cloud, whereas CFAR dynamically computes the threshold, making it more effective at suppressing such artifacts. In the walking scenario, where the human subject moves at a higher speed relative to the radar, all methods produce high proportion scores. MTI+CASO and MTI+Var, which employ MTI filters, show slight performance improvements over Mean+CASO and Mean+Var, which use mean removal for target detection. This suggests that MTI filtering is more effective in scenarios with greater relative movement, whereas the sitting scenario further supports that mean removal performs better when variations along the slow-time axis are minimal. Mean+CASO and MTI+CASO yield a higher total number of points because they apply the zoom-in spectrum after CFAR, resulting in a denser point cloud representation.

Table 5.21 presents the average reflection values from different body segments, as measured by each method.

Body Segment	Mean+CASO	MTI+CASO	MTI+Var	Mean + Var
Upper Torso	34.05%	31.57%	42.43%	44.13%
Lower Torso	45.19%	42.68%	19.86%	17.56%
Head	2.00%	2.80%	17.82%	18.70%

Continued on the next page

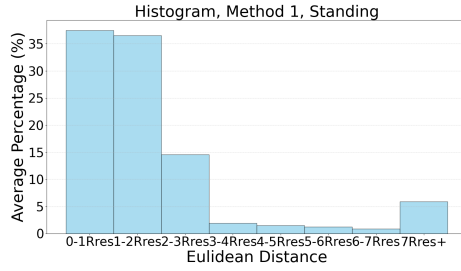
Continued from previous page

Body Segment	Mean+CASO	MTI+CASO	MTI+Var	Mean + Var
Arms	0.01%	0%	5.19%	4.34%
Legs	18.75%	22.94%	14.78%	15.27%

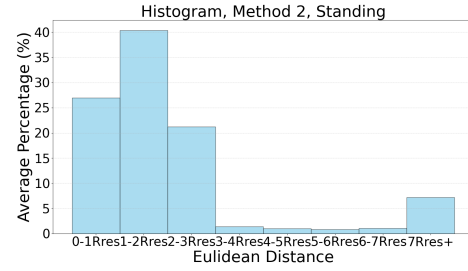
Table 5.21: Coverage for body segments

Mean+CASO and MTI+CASO exhibit comparable body segment coverage, whereas MTI+Var and Mean+ Var produce similar outcomes. Over 70% of the reflections in Mean+CASO and MTI+CASO originate from the torso, with the lower torso being the most reflective body segment. Reflections from the head and arms are minimal, while the remaining points correspond to the legs. MTI+Var and Mean+Var exhibit lower overall coverage of the torso. Unlike Mean+CASO and MTI+CASO, which primarily capture reflections from the lower torso, these methods show a higher concentration of points in the upper torso region. Another distinction is that MTI+VAR and Mean+VAR capture reflections from the head and arms, unlike Mean+CASO and MTI+CASO. This difference arises because the variance-based algorithm in MTI+VAR and Mean+VAR detects the nearest local maximum in each direction (θ, ϕ) of the spatial spectrum. This approach enables the selection of weaker reflection points that might otherwise be suppressed by stronger reflections from the torso within the same range bin when applying 2D CFAR across the azimuth-elevation spectrum.

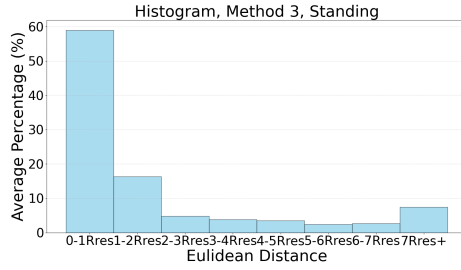
Figure 5.15 to Figure 5.19 illustrate the absolute distances of points from the ground-truth human mesh. A higher percentage of points within one range bin distance suggests that the variance-based method effectively captures radar reflections from the skin while filtering out points behind it. This is consistent with the fact that a 60 GHz mmWave signal can only penetrate 0.4 mm into human skin tissue [9]. However, the variance-based method is more prone to including artifacts, which appear as local maxima. As a result, MTI+VAR and Mean+VAR, which rely on this method, tend to produce more points that deviate significantly from the human mesh compared to Mean+CASO and MTI+CASO. On the other hand, the CFAR algorithm can better suppress points that exhibit larger deviations from the ground-truth human mesh by leveraging a noise window. This enables the adaptation of a higher dynamic threshold, improving the suppression of outliers.



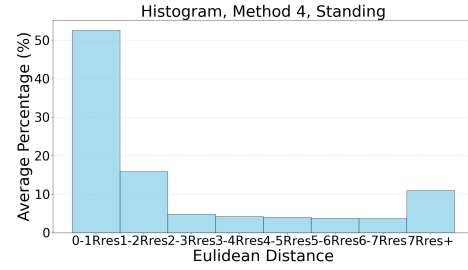
(a) Mean+CASO



(b) MTI+CASO

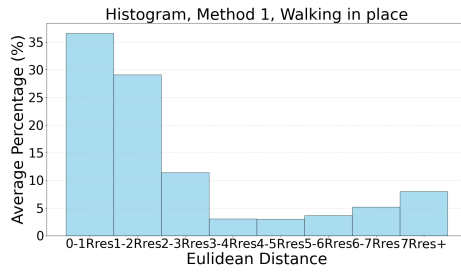


(c) MTI+Var

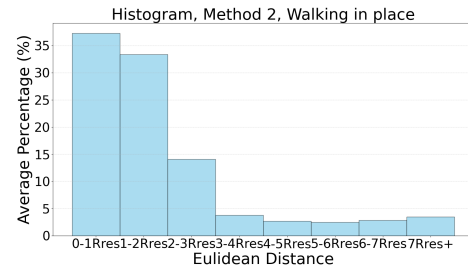


(d) Mean + Var

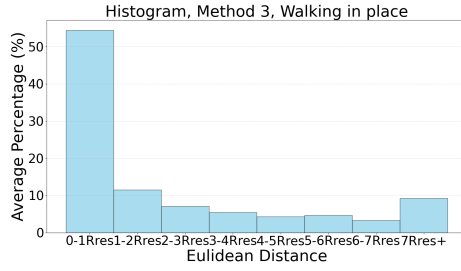
Figure 5.15: Histogram of Euclidean distances between points and the ground-truth human mesh in the Standing scenario



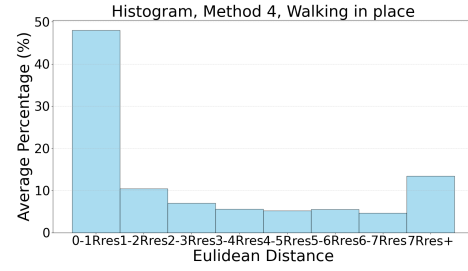
(a) Mean+CASO



(b) MTI+CASO

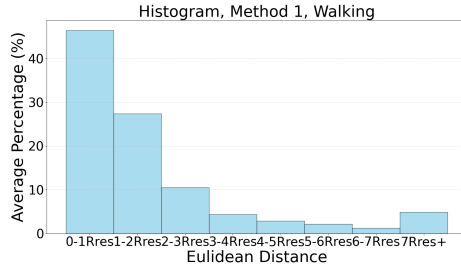


(c) MTI+Var

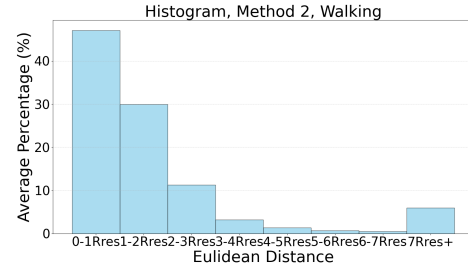


(d) Mean + Var

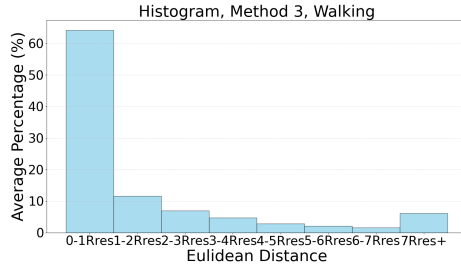
Figure 5.16: Histogram of Euclidean distances between points and the ground-truth human mesh for the Walking in place scenario



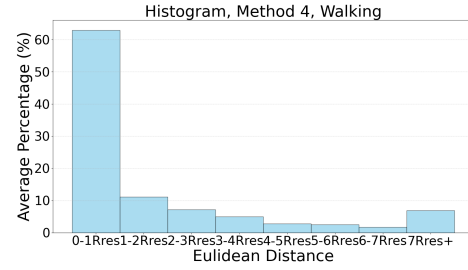
(a) Mean+CASO

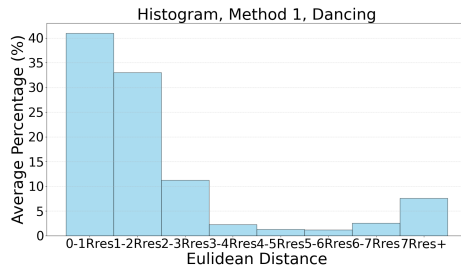


(b) MTI+CASO

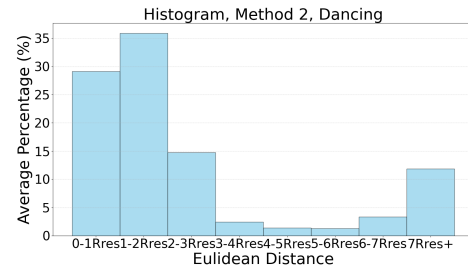


(c) MTI+Var

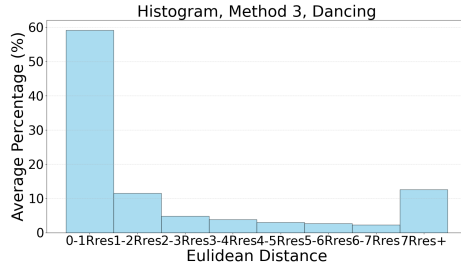




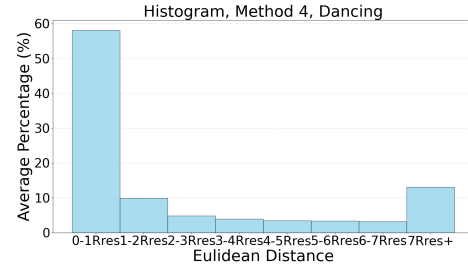
(a) Mean+CASO



(b) MTI+CASO



(c) MTI+Var



(d) Mean + Var

Figure 5.18: Histogram of Euclidean distances between points and the ground-truth human mesh in the Dancing scenario

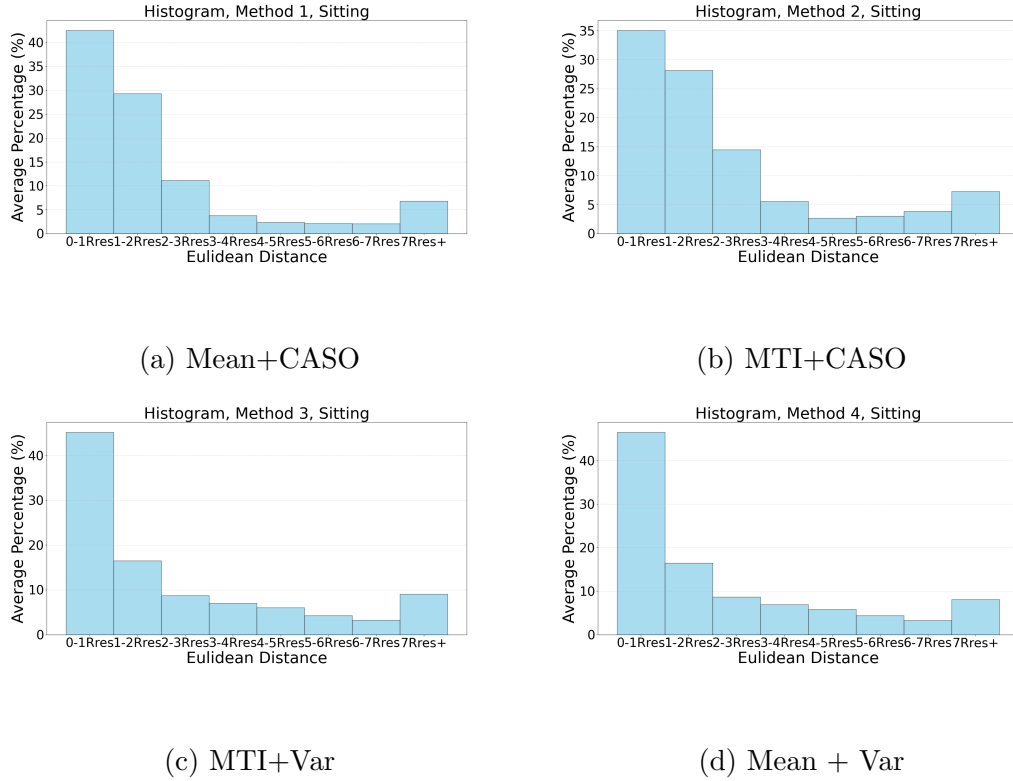


Figure 5.19: Histogram of Euclidean distances between points and the ground-truth human mesh in the Sitting scenario

5.4.2 Consistency Test

Table 5.22 and Table 5.23 present the average Hausdorff distances between one point cloud from a selected frame and point clouds from all other frames for standing and sitting scenarios, where the human subject remains stationary relative to the radar with minimal random body movements across all methods, Table 5.22 compares point clouds that include only points identified as reflections from the human body, while Table 5.23 compares point clouds containing all points. Table 5.24 and Table 5.25 demonstrate standard deviation of mean and maximum intensities across all frames for both standing and sitting scenarios. Table 5.26 and Table 5.28 show some statistics

of maximum intensities for standing and sitting scenarios across all frames. Table 5.27 and Table 5.29 demonstrate some statistics of mean intensities for standing and sitting scenarios across all frames.

Scenario	Mean+CASO	MTI+CASO	MTI+Var	Mean + Var
Standing	0.199	0.251	0.306	0.429
Sitting	0.196	0.323	0.442	0.427

Table 5.22: Hausdorff Distance of points classified as reflection from human body

Scenario	Mean+CASO	MTI+CASO	MTI+Var	Mean + Var
Standing	0.369	0.37	0.544	0.477
Sitting	0.369	0.574	0.588	0.57

Table 5.23: Hausdorff Distance of all points in point clouds

As shown in Table 5.22 and Table 5.23, MTI+Var and Mean+Var produce higher average Hausdorff distances for points identified as reflections from the human body as well as for the entire point cloud. This indicates that the point cloud exhibits lower consistency, even though the human subject remains in the same position across all frames. This finding further supports the conclusion from the coverage test, suggesting that variance-based methods are more prone to capturing reflections from body

segments with weaker signals. As a result, these methods are more susceptible to noise and artifacts.

MTI+Var and MTI+CASO, which use the MTI filter for clutter removal, both exhibit performance degradation, as shown in Table 5.19, in terms of the proportion of reflections and the Hausdorff distance evaluation when comparing the sitting scenario to the standing scenario. In contrast, for Mean+CASO and Mean+Var, which use mean removal, the differences in Hausdorff distances are less pronounced, both for the original point cloud and the filtered point cloud (containing only points identified as reflections). This is because mean removal better preserves small variations compared to the first-order MTI filter. As a result, smaller and localized random body movements are less suppressed, leading to fewer missed points and reduced spatial deviation. In contrast, mean removal retains these small variations, ensuring higher consistency across frames.

Scenario	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Standing	3.42	3.54	3.22	3.03
Sitting	7.46	6.96	6.72	6.63

Table 5.24: Standard Deviation of Mean Intensities across all frames

Scenario	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Standing	4.56	4.01	3.34	3.22
Sitting	9.16	8.21	7.01	6.88

Table 5.25: Standard Deviation of Maximum Intensities across all frames

Table 5.24 and Table 5.25 illustrate the standard deviation of the mean and maximum intensities across all frames for both standing and sitting scenarios. Table 5.26 to Table 5.29 demonstrate other related statistics. In general, the variance-based method has a lower standard deviation than the CFAR method. This is because the variance-based method consistently identifies the first local maximum exceeding a preset threshold, whereas CFAR uses a dynamic threshold, resulting in greater intensity variation. The standing scenario also has a lower standard deviation in mean and maximum intensities than the sitting scenario. This can be attributed to the effects of clutter removal techniques: when the subject is standing, their random body movements tend to be more consistent, resulting in reduced intensity variation after clutter removal. In contrast, during the sitting scenario, body movements are more subtle and localized, increasing the likelihood of human body reflections being suppressed, which in turn results in greater intensity variation.

	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Mean	145.61	151.9	149.32	143.02
Median	145.85	152.02	149.16	143.06
0.8 Quantile	149.07	155.29	152.53	145.62
0.95 Quantile	152.85	158.31	154.99	148.37

Table 5.26: Statistics for Maximum intensities of standing scenario

	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Mean	140.45	148.79	146.17	140.05
Median	140.5	149.09	145.91	140.04
0.8 Quantile	143.46	152.13	149.36	142.35
0.95 Quantile	145.66	154.31	151.62	145.22

Table 5.27: Statistics for Mean intensities of standing scenario

	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Mean	149.25	153.47	146.78	146.93

Continued on the next page

Continued from previous page

	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Median	149.25	154	146.4	146.76
0.8 Quantile	158.34	160.68	154.24	154.24
0.95 Quantile	163.79	165.59	158.59	158.59

Table 5.28: Statistics for Maximum intensities of sitting scenario

	Mean + CASO	MTI + CASO	MTI + Var	Mean + Var
Mean	144.34	149.72	142.98	143.05
Median	144.67	150.45	141.95	142.28
0.8 Quantile	151.93	155.62	150.29	150.32
0.95 Quantile	155.94	161.4	155.17	154.99

Table 5.29: Statistics for Mean intensities of sitting scenario

Chapter 6

Conclusion

To conclude, this thesis presents quantifiable evaluation metrics for both static targets and a human subject. For static target scenarios, the proposed metric includes Euclidean distance error, range error, azimuth angle error, and elevation angle error, providing a more comprehensive assessment of pipeline performance than Euclidean distance error alone. For human target scenarios, this work introduces coverage and consistency evaluations. The coverage test utilizes the Euclidean distance to the human mesh surface to determine the percentage of points that originate from the human body, while also assessing coverage across different body regions. Consistency evaluation is conducted under the condition that the human subject maintains a fixed position and posture across all frames. This involves computing the mean and maximum intensity values over all frames and analyzing the Hausdorff distance between the reference frame's point cloud and those of all other frames. These metrics offer a detailed quantitative assessment of the generated point clouds and the effectiveness of the underlying processing pipelines.

The experiment demonstrates that angle estimation accuracy decreases with increasing range and incident angles. For the static target scenario, variance-based target detection combined with MVDR spectrum estimation yields the best performance. The CFAR algorithm performs less effectively than the variance-based method. Among the CFAR variants, CFAR-CASO shows slightly superior performance compared to CFAR-CA in most trials. The MUSIC algorithm performs similarly in the single-target test but experiences a decline in performance in the three-target test. In the human subject tests, the CFAR algorithm requires larger noise windows and guard cells than in the static target tests. The optimal configuration for the CFAR algorithm in the human subject tests includes a noise window size of 14 and a guard cell dimension of 8 for both 1D and 2D CFAR. However, configurations with larger noise windows and guard cell dimensions deliver comparable results. All pipelines in the human subject test employed MVDR spatial spectrum estimation. The pipeline achieving the best overall performance combines mean removal for clutter removal and CFAR with a zoom-in spectrum for target detection. Pipelines that use the variance-based method for target detection exhibit poorer consistency compared to CFAR-based methods in the human target tests.

6.1 Limitations and Future Work

This study was conducted using a single radar device and configuration, without an in-depth analysis of how variations in range resolution, angular resolution, and the number of chirps affect pipeline performance for each metric discussed in this thesis. The experiment was carried out in a controlled laboratory setting, lacking diversity in human subject scenarios. Future research will focus on collecting a more diverse

dataset, incorporating human subjects with varying body shapes. Additionally, real-world scenarios will be included in data collection, and a broader range of radar configurations and devices will be explored.

This study evaluates a limited set of commonly used algorithms and a standardized pipeline based on previous research, focusing primarily on target detection for human subjects and static point targets. Future research could explore a broader range of pipelines and algorithms, incorporating both existing and newly developed evaluation metrics. Additionally, future work will aim to establish a stronger correlation between point cloud evaluation metrics and downstream task performance, analyzing how high-scoring point clouds impact their respective applications. Further refinements in evaluation metrics and validation across diverse radar configurations and environmental conditions will enhance their robustness and applicability.

Bibliography

- [1] M. Ash, M. Ritchie, and K. Chetty. On the application of digital moving target indication techniques to short-range FMCW radar data. *IEEE Sensors Journal*, 18(10):4167–4175, 2018. doi: 10.1109/JSEN.2018.2823588.
- [2] G. Brooker. Understanding millimetre wave FMCW radars. *1st International Conference on Sensing Technology*, 01 2005.
- [3] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969. doi: 10.1109/PROC.1969.7278.
- [4] J. Cozzens and M. Sousa. Source enumeration in a correlated signal environment. *IEEE Transactions on Signal Processing*, 42(2):304–317, 1994. doi: 10.1109/78.275604.
- [5] H. Cui, S. Zhong, J. Wu, Z. Shen, N. Dahnoun, and Y. Zhao. Milipoint: a point cloud dataset for mmwave radar. In *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23*, Red Hook, NY, USA, 2023. Curran Associates Inc.
- [6] Y. Feng, J. Zhao, C. Wang, L. Xie, and S. Lu. 3d bounding box estimation based on cots mmwave radar via moving scanning. *Proc. ACM Interact. Mob.*

- Wearable Ubiquitous Technol.*, 8(4), Nov. 2024. doi: 10.1145/3699758. URL <https://doi.org/10.1145/3699758>.
- [7] T. Gu, Z. Fang, Z. Yang, P. Hu, and P. Mohapatra. mmsense: Multi-person detection and identification via mmwave sensing. In *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, mmNets '19, page 45–50, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369329. doi: 10.1145/3349624.3356765. URL <https://doi.org/10.1145/3349624.3356765>.
- [8] S. Gupta, P. K. Rai, A. Kumar, P. K. Yalavarthy, and L. R. Cenkeramaddi. Target classification by mmWave FMCW radars using machine learning on range-angle images. *IEEE Sensors Journal*, 21(18):19993–20001, 2021. doi: 10.1109/JSEN.2021.3092583.
- [9] I. T. M. H. Hikage T, Ozaki R. Novel 60 ghz band spatial synthetic exposure setup to investigate biological effects of 5g and beyond wireless systems on human body. *Front Public Health*, 2021. doi: 10.1109/JSEN.2018.2823588.
- [10] K. Hu, H. Liao, M. Li, and F. Wang. Mmcount: Stationary crowd counting system based on commodity millimeter-wave radar. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 56–60, 2024. doi: 10.1109/ICASSP48485.2024.10447408.
- [11] T. Instruments. DCA1000EVM, 2019. URL <https://www.ti.com/tool/DCA1000EVM>.
- [12] T. Instruments. 3D People Counting Demo Software Implementation Guide,

- 08 2022. URL https://dev.ti.com/tirex/explore/node?a=1AslXXD__1.00.00.26&node=A__ANL8AwYmp.Qo2i.UiR1E3A__radar_toolbox__1AslXXD__1.00.00.26.
- [13] T. Instruments. Iwr6843, 2022. URL <https://www.ti.com/product/IWR6843>.
- [14] F. Jin, A. Sengupta, and S. Cao. mmfall: Fall detection using 4-d mmwave radar and a hybrid variational rnn autoencoder. *IEEE Transactions on Automation Science and Engineering*, 19(2):1245–1257, 2022. doi: 10.1109/TASE.2020.3042158.
- [15] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, Oct. 2015.
- [16] S. Lovescu, C.; Rao. The fundamentals of millimeter wave radar sensors, 2021. URL <https://www.ti.com/lit/pdf/spyy005>.
- [17] NaturalPoint. OptiTrack motive motion capture system, 2024. URL <https://www.optitrack.com/>.
- [18] A. Ngo, M. Bauer, and M. Resch. Deep evaluation metric: Learning to evaluate simulated radar point clouds for virtual testing of autonomous driving, 04 2021.
- [19] L. Ren, A. G. Yarovoy, and F. Fioranelli. Grouped people counting using mm-wave fmcw mimo radar. *IEEE Internet of Things Journal*, 10(22):20107–20119, 2023. doi: 10.1109/JIOT.2023.3282797.
- [20] N. Scheiner, O. Schumann, F. Kraus, N. Appenrodt, J. Dickmann, and B. Sick. Off-the-shelf sensor vs. experimental radar - how much resolution is necessary in

- automotive radar classification? In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pages 1–8, 2020. doi: 10.23919/FUSION45008.2020.9190338.
- [21] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986. doi: 10.1109/TAP.1986.1143830.
- [22] A. Sengupta, F. Jin, R. Zhang, and S. Cao. mm-pose: Real-time human skeletal posture estimation using mmWave radars and cnns. *IEEE Sensors Journal*, PP: 1–1, 05 2020. doi: 10.1109/JSEN.2020.2991741.
- [23] T.-J. Shan, M. Wax, and T. Kailath. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):806–811, 1985. doi: 10.1109/TASSP.1985.1164649.
- [24] Z. Shen, J. Nunez-Yanez, and N. Dahnoun. Advanced millimeter-wave radar system for real-time multiple-human tracking and fall detection. *Sensors*, 24: 3660, 06 2024. doi: 10.3390/s24113660.
- [25] A. A. Taha and A. Hanbury. An efficient algorithm for calculating the exact hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 03 2015. doi: 10.1109/TPAMI.2015.2408351.
- [26] B. Van Veen and K. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, 1988. doi: 10.1109/53.665.
- [27] M. Wax and T. Kailath. Detection of signals by information theoretic criteria.

- IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985. doi: 10.1109/TASSP.1985.1164557.
- [28] C. Will, P. Vaishnav, A. Chakraborty, and A. Santra. Human target detection, tracking, and classification using 24-ghz FMCW radar. *IEEE Sensors Journal*, 19(17):7283–7299, 2019. doi: 10.1109/JSEN.2019.2914365.
- [29] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su. mmMesh: towards 3d real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys ’21, page 269–282, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384438. doi: 10.1145/3458864.3467679. URL <https://doi.org/10.1145/3458864.3467679>.
- [30] B. Zhang, B. Jiang, R. Zheng, X. Zhang, J. Li, and Q. Xu. Pi-vimo: Physiology-inspired robust vital sign monitoring using mmWave radars. *ACM Transactions on Internet of Things*, 4, 03 2023. doi: 10.1145/3589347.
- [31] B. Zhang, Z. Zhou, B. Jiang, and R. Zheng. Super: Seated upper body pose estimation using mmwave radars. In *2024 IEEE/ACM Ninth International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 181–191, 2024. doi: 10.1109/IoTDI61053.2024.00020.
- [32] D. Zhang, X. Zhang, S. Li, Y. Xie, Y. Li, X. Wang, and D. Zhang. Lt-fall: The design and implementation of a life-threatening fall detection and alarming system. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(1), Mar. 2023. doi: 10.1145/3580835. URL <https://doi.org/10.1145/3580835>.

- [33] F. Zhang, C. Wu, B. Wang, and K. J. R. Liu. mmeye: Super-resolution millimeter wave imaging. *IEEE Internet of Things Journal*, 8(8):6995–7008, 2021. doi: 10.1109/JIOT.2020.3037836.